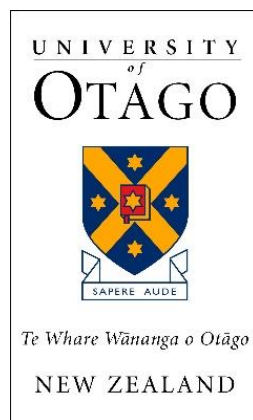


How to Think About Enhancement: A Welfarist Approach



Hazem Zohny

Submitted for the degree of
Doctor of Philosophy

Bioethics Centre
The University of Otago

December 2016

Abstract

This thesis sets out to answer two questions: What are the implications of the welfarist account of enhancement, and on what basis should enhancements be developed and distributed? The welfarist account defines enhancement as any change in the biology or psychology of a person which increases their chances of leading a good life in the relevant set of circumstances. Herein, I systematically break down the account, interrogating each component of its definition, and explore its scope and limitations. My goal is to show how, despite some of its counter-intuitive implications, the account can help us think more clearly about enhancement, allowing opponents and proponents of enhancement to articulate their arguments in ways that offer a constructive way forward in debating this topic. It does so by avoiding what I will argue are contestable terminologies, boundaries and controversial values, instead re-orienting our attention to what is of central value to questions of enhancement: how bodily and mental states tend to impact well-being. On this account, for something to be an enhancement, it must tend to improve people's well-being. In that way, the account compels us to first be clear about what we mean by a good life and what sorts of bodily and mental changes contribute to it, before we consider an intervention to be enhancing. However, this by no means resolves all ethical concerns: some interventions could be good for users – that is, they may be enhancements – but their development or distribution may nevertheless be unjust. This relates to the second question driving this thesis, and here I argue that the welfare-maximizing approach of utilitarianism can act to adjudicate between different moral stances on the development and distribution of enhancements. Tying it all together, I propose and defend a mechanism, the Well-Being Impact Fund, that links the welfarist account with a utilitarian approach in a way that reduces the chances of emerging enhancements leading to a neglect of vital disease treatments, or an exacerbation of existing inequalities within and between societies. Altogether, I argue this thesis provides a framework for thinking about enhancement in a way that can largely overcome theoretical disagreements, and that offers actionable conclusions and strategies for responding to challenges raised by enhancement.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisors Mike King and John McMillan for their continuous support of my PhD study and related research, but also for their patience, motivation and always insightful (and often near instantaneous!) feedback. I could not have imagined having better supervisors for my PhD. As my primary supervisor, Mike King has been a constant ally who masterfully managed that subtle combination of challenging me, guiding me, but also giving me my own space.

My sincere thanks to Lynley Anderson for being a supportive and helpful convener during my annual reviews, and to the Bioethics Centre's Postgraduate Forum for letting me throw around my often patchy ideas. I would also like to thank The University of Otago for funding my PhD study – this simply would not have been possible with its financial support.

Finally, I thank the audience that attended a talk I gave at the Oxford Uehiro Centre for Practical Ethics, in particular Tom Douglas, Julian Savulescu and Walter Sinnott-Armstrong. Their feedback and discussions about my thesis in general have been particularly insightful.

Contents

Abstract	I
Acknowledgements	II
Introduction	1
Plan of the thesis	8
Chapter 1 Preliminaries	12
1. Enhancement and the risk of speculative ethics	13
2. Methodological guidelines	18
Chapter 2 Enhancement and Well-Being	22
1. The welfarist account and its critics	23
2. A deeper understanding of the welfarist account	25
3. Objections to the welfarist account	27
3.1 Renders “enhancement” too broadly	27
3.2 Fails to distinguish therapy from enhancement	30
3.3 Overly dependent on the contested concept of well-being	31
3.4 Does not offer timely evaluations	34
3.5 Sidesteps the ethical debate on enhancement	35
3.6 Underestimates social forces	37
4. Enhancement, disability, and the riddle of the relevant circumstances	39
4.1 The conceptual link between enhancement and disability	42
4.2 The mislabelling objection	45
4.3 Excluding social prejudice (and why this fails)	47
4.4 An alternative to excluding social prejudice	51
5. Conclusion and further thoughts	54
Chapter 3 Enhancing the Extended Mind	56
1. Where do mental states happen?	57
2. An objection to the extended mind thesis	61
3. The implication of the extended mind thesis for cognitive enhancement	64
4. Conclusion	69
Chapter 4 Welfarist Moral Enhancement	71
1. The welfarist account of moral enhancement	72
2. Cognitive enhancement as indirect moral enhancement	82
3 Evidence for the relationship between cognition and moral dispositions	86

3.1 A moral Flynn effect	86
3.2 Intelligence and cooperation	88
3.3 Intelligence and being more understanding	89
3.4 Intelligence and being more liberal.....	90
3.5 Shifts in morality: cultural or cognitive?	91
4. Moral judgments versus moral motivation.....	94
4.1 Moral motivation enhancement	96
5. Conclusion	98
Chapter 5 Smart Drugs and Welfarist Enhancement	100
1. Recurring presumptions about smart drugs	101
2. Cognitive enhancement in the literature	102
3. Four presumptions	104
3.1 Presumption #1: Cognitive enhancement drugs already exist	104
3.2 Presumption #2: Cognitive enhancement is separable from non-cognitive enhancement	106
3.3 Presumption #3: There are ethical issues that are uniquely raised by cognitive enhancements.....	110
3.4 Presumption #4: There is evidence that PCE is increasingly in use	112
4. The cost of hype	116
5. A positive shift.....	118
6. Getting clearer about cognitive enhancement	118
7. Are smart drugs welfarist enhancements?	120
7.1 The welfarist account and group enhancement	125
Chapter 6 Enhancements and Maximizing Well-Being Impartially	129
1. Utilitarianism's building blocks	131
1.1 Welfarism	132
1.2 Consequentialism	134
1.3 Sum ranking.....	135
2. Utilitarianism and enhancement.....	137
3. The principle of diminishing marginal utility.....	142
3.1 The wealthitarian fallacy	145
4. Objections to Utilitarianism	147
4.1 The distributive objection	147
4.2 Do human capacities exhibit diminishing marginal utility?	153
4.3 Utilitarianism and people with impairments	157
5. Utilitarianism, the state, and enhancement	161

6. Conclusion	165
Chapter 7 Enhancement & Inequality: The Well-Being Impact Fund	166
1. Enhancement and inequality	167
2. The Global Institute for Justice in Innovation	168
3. The Health Impact Fund	173
4. What about transmedical enhancements?	177
4.1 Incorporating the welfarist account of enhancement	178
5. The Well-Being Impact Fund	180
5.1 The Well-Being Impact Fund in action	181
6. Five objections.....	185
7. Conclusion	189
Conclusion	190
Bibliography	198

Introduction

How should we understand the concept of enhancement? And on what basis should we promote or restrict its pursuit? My goal for this thesis is to answer these two questions. Specifically, I aim to refine and defend an account of enhancement that characterizes it as a biological or psychological change that tends to improve a person's well-being (Savulescu, Sandberg, & Kahane, 2011). This welfarist account, I will argue, holds the promise of clarifying our thinking about enhancement, allowing us to more pragmatically deliberate about and respond to enhancement technologies in a way that is useful for practical ethicists and policy makers. My subsequent goal is to use this account to formulate a framework for thinking about the promotion or restriction of enhancement, and specifically, to respond to the concern that enhancements may exacerbate inequalities within and between societies, or lead to a greater neglect of disease treatment.

As such, my two driving questions for this thesis are:

- What is the welfarist account of enhancement, and what are its full implications?
- On what basis should enhancements (in the welfarist sense) be developed and distributed?

These are interrelated questions: we are unlikely to agree about answers to the second question without first having a clear, shared response to the first one – what is it exactly are we debating whether to promote or restrict?

It is worth highlighting this relationship between the two questions from the outset. Consider an approach that shuns this premise and homes in on a question of the second variety. Leon Kass (2003) offers a prime example: “Needless arguments about whether or not something is or is not an ‘enhancement’ get in the way of the proper question: What are the good and bad uses of biotechnical power?”

This is an appealing approach to the extent that it focuses on what seems most pressing: real-world ethical issues, not abstruse conceptual analyses. And yet, in avoiding one question (what is enhancement?), it raises others: what constitutes “biotechnical power” exactly? Does it include the ancient processes of domesticating animals and cultivating plants for human purposes, and if not, why not? Moreover, why presume that only this specific type of power

should be the sole concern of this topic? Do non-biotechnical forms of power raise fundamentally different ethical issues?

What this shows is that avoiding a definition of enhancement only raises more questions and ambiguities, potentially priming opponents and proponents of enhancement to speak past each other. This is because how we understand something can have major consequences for how we view its ethical implications, and therefore how we regulate it. For instance, whether we understand a foetus as a soul-endowed person or a bundle of cells will have major implications for what we think we should be allowed to do with them (abort them, destroy them for stem cells, and so on). Similarly, whether we understand enhancements as interventions that, say, entail a loss of the features that make life valuable, or as ones that increase what is valuable about life, will have significant implications for how we regulate their use and development. Because of that, we need a coherent, shared understanding of what enhancement is before we can effectively debate what we should do about its promotion or restriction. Pellegrino (2004) conveys this point uncompromisingly: “No boundary between morally valid and invalid uses of biotechnology can be established without at least a working definition [of enhancement].”

To that end, what is enhancement? Definitions of the term typically assume that it is distinct from therapy, augmenting or improving our capacities beyond what is required for normal or healthy functioning.¹ In that way, enhancements do not respond to genuine medical needs, whether those needs are defined in terms of disease, impairment, deviations from species-typical functioning, or by reference to prevailing medical understandings (DeGrazia, 2005).

However, from the perspective of the welfarist account, this is an unhelpful premise from which to proceed. As many have pointed out (e.g. Bostrom & Roache, 2008; Daniels, 2000; Harris & Chan, 2008) the boundary between what is therapeutic and enhancing is blurry at best. Standard contemporary medicine now encompasses many techniques that go beyond merely restoring function. These include preventative medicine, plastic surgery, contraceptive devices, and palliative care (Bostrom & Roache, 2008). Vaccination is an example of a standard contemporary medical intervention. On the one hand, it appears to be a preventative medical measure, allowing us to maintain healthy functioning. Yet it also seems to be an enhancement, fortifying the immune system beyond how it would normally function. There are other examples: contraceptive pills do not (ordinarily) restore health or prevent disease – instead, they augment

¹ By capacities I mean our abilities to perform or produce things, whether using our bodies or mental faculties. By functioning I mean the performance of something in terms of what it is normally used for: the function of the heart is to pump blood and its healthy functioning entails doing so effectively. I use functioning and capacities interchangeably.

women's control over their reproductive lives in ways that would not be otherwise possible. Alternatively, an artificial limb or organ may be used to restore a lost capacity or function, but the increasing sophistication and relative invulnerability of such prostheses may also make them functionally enhancing.

A number of scholars acknowledge these grey areas, but still insist that the distinction can be made in most cases (Daniels, 2000; DeGrazia, 2005). However, what is the distinction useful for? Is it that, by definition, therapy is permissible but enhancement is not? Clearly no: there may be many interventions that restore health that are not obviously permissible (such as restoring dysfunctional organs by harvesting human-pig chimeras or, more hypothetically, harvesting non-conscious humans specifically bred for their organs). Is it that therapy should be publicly funded but not enhancement? Again, no: it is not plausible to fund everything that is deemed therapeutic, while some interventions with enhancing elements (like prostheses and contraceptive pills) are, and arguably should be, funded. Is it that we should develop more and better therapies, but not enhancements? Once again, this is not an obvious imperative. Like the previous questions, it requires us to question what the end goal of therapy is – is it simply to restore function, or is restored function merely a means to some other end?

The welfarist account stems from a plausible premise that both therapy and enhancement are, generally, pursued by individuals for the same ultimate end – which is to improve their well-being. By that I mean they are pursued in the hope of adding to whatever it is that constitutes or contributes to the good in their life. Given this shared end goal, the welfarist approach avoids entangling the concept of enhancement (and therapy, for that matter) with contestable boundaries tied to problematic concepts like “normal” and “abnormal” traits and functioning, along with unproductive attempts to distinguish them. In fact, it makes no distinction between enhancement and therapy at all. Instead, it directs our deliberations down what I think is a more fruitful avenue of thought. Rather than using a possibly arbitrary boundary between enhancement and therapy to delineate what is permissible, or what should be funded or developed, it directs our thinking to the relationship between a given bodily or mental intervention, and its likely effect on the well-being of the person who undergoes it.

For a brief illustration of how it may do this, consider a child who is projected to be extremely short due to an otherwise benign brain tumor stunting their growth, and another child who will be equally short due to genes inherited from short parents (Daniels, 2000). In other words, one has an identifiable dysfunction (a brain tumor) while the other does not (having “short genes”

is not ordinarily thought of as a dysfunction).² Does one therefore require therapy while the other can only hope for enhancement? And does one therefore deserve public funding for such an intervention – which would be the case in the US if the child is given human growth hormones (Daniels, 2000) – while the other does not? The welfarist account asks us to skip these irrelevant distinctions and focus on whether an intervention that would change their stature would likely improve their well-being, and further, whether we should strive to improve their well-being in that specific way. It is these questions that should occupy our deliberations instead.

In that way, it is an appealing approach to the extent that it reorients our attention to what seems of central value here: how bodily and mental states impact well-being, and what we should do about those states. It is the centrality of well-being to these questions that I think makes this approach a more useful account for guiding our thinking about enhancement. The assumption here is not that well-being is all that matters, but that it is the kind of universal value that can act as a common currency for articulating different stances on the development or use of bodily and mental interventions. Hence, while well-being is not necessarily considered as the sole ultimate value, it can nevertheless be used to cash out other values (something that I will highlight in chapter five). In other words, I use well-being here as a pragmatic metric of value against which other values can be measured, one that can help us reach principled compromises when these other values conflict (Greene, 2014).

To that end, much of this thesis elaborates on the welfarist account, exploring its implications and limitations. Currently, a number of gaps remain in the account. For instance, while what constitutes the biology of a person is clear enough, what constitutes a person's psychology? Embedded and extended theories of the mind may have crucial implications for the scope of this account, and these have not previously been considered. Similarly, while the account is presented as a general approach to enhancement (Savulescu et al., 2011), it is not clear, and its formulators have not specified, how it can account for moral enhancement. If an enhancement is something that tends to increase our well-being, must all moral enhancements then be good for us, in a prudential sense? What about interventions that make us kinder but not obviously better off? It's not clear the welfarist account can accommodate such interventions, which are commonly regarded as a form of enhancement. More broadly, how useful is this account as a guide for policy-making? It may be a philosophically useful account, but is it relevant to how we ought to think about the regulation of emerging technologies? Indeed, what is a welfarist about

² Of course, what exactly qualifies as a dysfunction is a further matter of contention. It is not clear that the study of taxonomy, genetics or physiology can describe what a natural state of functioning is, nor what "natural" refers to more generally (Ereshefsky, 2009).

enhancement to make of so-called smart drugs, which occupy much of the discussion in the enhancement literature?

These are gaps in the literature that have not been adequately investigated. I intend to fill these gaps in the account. In doing so, I clarify and polish it, adding and responding to the criticisms it has amassed over the past five years. To reiterate, the importance of doing this work is straightforward: if we are to tackle the challenges posed by certain existing and emerging technologies relevant to enhancement, we need a shared, coherent and useful understanding of the term (Pellegrino, 2004; Savulescu et al., 2011). The welfarist account holds that promise, but it needs work.

Hence my first question above: What is the welfarist account of enhancement, and what are its full implications? As I noted, my second goal is to investigate how a welfarist approach can also help us resolve broader ethical questions about enhancements, namely issues related to distributive justice. This is crucial because, even if we can agree about what makes an intervention good for a person's well-being, this tells us nothing about who should have access to it and whether it should be developed or made accessible. My goal, then, is not only to further develop the welfarist account as a useful framework for thinking about enhancement, but to use that account to draw a specific proposal about how we ought to develop and distribute enhancements, both within societies and between them. Specifically, it is a proposal to mitigate what I believe is the most serious problem raised by enhancement, which is that it may exacerbate existing inequalities. Hence, to repeat the second question driving this thesis: on what basis should enhancements (in the welfarist sense) be developed and distributed?

Note, however, the broad phrasing implied by "on what basis", as opposed to *what specific regulations* should determine how we develop and distribute enhancements. The reason for this is because the enhancements that occupy much of the literature are either only emerging or hypothetical. As we will see, we do not yet have the necessary details to formulate specific regulations for them. Moreover, there can be no monolithic answer to policy issues, unless it is so vague as to be otiose, such as by responding "reasonably" when asked "how should we regulate enhancements?" Unpacking what "reasonably" entails means having access to the details and socio-political circumstances in which the technology is used, which is not the case for emerging or hypothetical technologies. Hence, my goal is to provide a pragmatic framework for approaching questions of innovation and distribution as they pertain to enhancement. This framework is, in a sense, an attempt to deploy the welfarist account of enhancement using an ethical theory. This theory springs naturally from my framework for understanding

enhancement: it too is welfarist, and more specifically, it is impartially welfare-maximizing – that is, it is utilitarian.

My appeal to utilitarianism later in this thesis will not be based on any claim that it is necessarily the correct ethical theory – we need not even presume that a correct ethical theory exists. Instead, like the welfarist account of enhancement, utilitarianism highlights a common human value – well-being – and appeals to its promotion in trying to resolve disputes about conflicting interests and values. In the same way that this shared value of well-being can help us resolve and better understand disputes about interventions that modify our bodies or mind, that same value, I will argue, can help us resolve and better understand disputes about what to do about their development and distribution. Moreover, because almost all ethical theories join commonsense morality in placing moral weight on preserving and promoting well-being, while also considering consequences as relevant to the moral assessment of actions and policies (Shaw, 2016), appreciating what utilitarianism has to say about enhancement is something that anyone interested in the ethics of enhancement should find relevant and insightful.

Before proceeding with an outline of my thesis, it is worth acknowledging at this early stage that the welfarist account has been framed as being favoured by proponents of enhancement (Giubilini & Sanyal, 2016), potentially biasing the debate against opponents, who typically seek to maintain the moral relevance of the therapy-enhancement distinction. This is because, in part at least, maintaining it allows opponents – sometimes labelled “bio-conservatives” – to easily name the class of interventions they oppose. If this is true, then one might argue the welfarist account is not well-suited to act as the useful framework I hope it can be, clarifying and helping resolve disputes over the permissibility of enhancements.

However, it is not clear how the therapy-enhancement distinction helps bio-conservatives name the class of interventions they oppose. By arguing against any and all interventions that do not respond to genuine medical needs, a proponent of enhancement need only identify one example of enhancement where it is implausible to justify outright opposition to it (such as vaccination or contraceptive pills). For the reasonable bio-conservative, this would mean that they then need to further specify what enhancements, aside from those uncontroversial ones, they are specifically referring to, or why those interventions identified by the proponent are not actually enhancements but therapies (or some other intervention entirely). Doing this extra work would entail either sharpening the therapy-enhancement distinction in a likely arbitrary way, or elaborating on what a genuine medical need is. Either way, merely maintaining the

distinction does not automatically help bio-conservatives name what precise class of interventions they are against.

Secondly, in-principle objections to enhancements (as contrasted to therapies) tend to be rooted in a belief that enhancements would be bad for our chances of leading good lives. Consider the claim that enhancement may deprive us of authenticity (Sandel, 2007), or that it may relinquish us of the hardships that give value and meaning to our lives (Mehlman, 2012), or that it may undermine our dignity (Fukuyama, 2002; Kass, 2002), or that genetic manipulations of offspring may rob them of autonomy (Habermas, 2003). At least in a policy context, how else are we to make sense of the weight of these concerns if not for the fact that, if they are correct, they signify a diminished or complete loss of an ability to lead good, worthwhile lives? True, while some of their weight may derive from other, non-welfarist concerns, it is clear that at least some significant portion of their importance as objections is rooted in a concern for our well-being.

As such, a welfarist understanding of enhancement can in fact arm bio-conservatives with an effective point of focus for their arguments. As they stand, their opponents have been shutting down their arguments by questioning their secularity (Kahane, 2011; Pinker, 2008), the relevance of appealing to nature and the conceptual confusions caused by doing so (Buchanan, 2011; Nielsen, 2011), as well as the value of rooting their arguments in contentious, if not “useless”, terms like dignity (Macklin, 2003; Pinker, 2008). However, once these bio-conservative concerns are voiced in welfarist terms, it can become very clear if not self-evident what exactly is at stake.

Of course, well-being itself is a contentious concept. However, what is not contentious about it is that it is a commonly accepted and important human value: while we may disagree about what it is, we all agree that any plausible ethical theory places much weight on it. Ultimately, as we will see, the welfarist account may uncover that what bio-conservatives and their opponents actually disagree about is what constitutes a good life, not whether a certain intervention is permissible or not. If so, a welfarist account of enhancement may simply entail that we need to move much of the discussion upstream, focusing on what it is that constitutes a good life to begin with, and then deliberating about how a bodily or psychological intervention is likely to contribute to such a life.

Plan of the thesis

Chapter 1

The first chapter highlights my methodological approach to this thesis. In particular, I focus on the issue of speculative ethics: given uncertainties about how and whether enhancement technologies will arise, and given far more pressing and existing moral problems currently facing the world, why squander my efforts on this arguably speculative topic? It is an important question in so far as it challenges the rationale of my thesis. Here, I articulate the nature of the concerns about speculative ethics and respond to them. In doing so, I set out certain key methodological ideals that I intend to follow throughout the thesis.

Chapter 2

This chapter explains the welfarist account of enhancement in detail, elaborating on its implications and responding to its most significant criticisms. I argue that the objections raised against it are either unfounded or addressable, but I also further critically explore its implications by examining how the account relates to the concept of disability. A key point in this chapter will be that the welfarist account expands the scope of enhancement as normally understood, which makes the conclusions reached in this thesis not only relevant to emerging, augmentative interventions affecting the body and mind, but also to therapies in general, as well as non-science or technology-based interventions.

Chapter 3

Enhancements tend to be characterized as limited to internal changes to a person that take place in their biology or psychology. However, while what is entailed by a person's biology is clear enough, what constitutes a psychology is not. We tend to think of the skull as the relevant boundary within which our psychologies reside or unfold. But, as I will argue here, this view runs into some difficulty when we examine the arguments for the so-called extended mind thesis (Clark & Chalmers, 1998). This is the thesis that the correct understanding of cognition entails that it can, in the right circumstances, literally take place outside the head in processes that entangle it with external props and devices.

If this view is correct, it further broadens the scope of the welfarist account: if innovations like smart phones can literally entail changes to our psychology when parts of our cognitive processes integrate with them, then they may constitute enhancements in ways that are not fundamentally different from internal changes to our body or brain. This, I will argue, ought to

significantly impact how we view cognitive enhancement. Cognitive enhancement is ubiquitous and we have been engaging in it since the invention of the very first tools.

Chapter 4

With the most significant implications the welfarist account laid out, I then explore a potential limitation in chapter four – specifically, can the welfarist account makes sense of moral enhancement? If enhancements are only things that tend to make users’ lives go better, and becoming more moral can often entail sacrificing our own interests, there may be a conflict between the welfarist account (given its self-regarding focus) and moral enhancement. Here I examine this conflict and argue for an understanding of moral enhancement as a form of cognitive (welfarist) enhancement. To make this case, I evaluate the evidence for the relationship between certain cognitive abilities and moral dispositions.

Chapter 5

At this stage I move to further consider the implications and limitations of the welfarist account by deploying it: will it provide the kind of approach that practical ethicists and policy makers can use in deliberating about some interventions? In this chapter, I explore this question in the context of the debate on so-called smart drugs such as modafinil and Adderall. I focus on this topic specifically (as opposed to genetic or some other type of enhancement) for two reasons: it has arguably engendered the most discussion in the enhancement literature, and it is a prime example of how poor conceptual clarity, inadequate reliance on empirical data, and a propensity for hype get in the way of clear thinking and sound policy recommendations.

Here, I show how the welfarist account requires us to investigate the evidence for these drugs’ efficacy, as well as the relationship between improved cognition and well-being. As I will argue, we may have to slightly alter the scope of the welfarist account to make it better suited for policy purposes in general.

Chapter 6

In chapter six, I begin answering the second question driving this thesis: on what basis should enhancements be developed and distributed? At this stage, it will be clear that we are talking about a class of interventions and innovations that are much broader than we may have initially expected: these include therapies as well as certain devices, and we are no longer talking about possible future innovations, but current ones as well. We will also understand the relationship between cognition and moral motives and behavior, which may impact how we ought to develop and distribute these innovations. Here, in the same way I have argued that well-being

can act as a common currency for cashing out other values in helping us thinking more clearly about enhancement, I will argue that that this same value can also help resolve questions about the development and distribution of enhancement; specifically, I will defend a welfare-maximizing approach to distributive issues that enhancements arises. This is utilitarianism, and in this chapter I respond to some objections to this theory and show how it would interact with the welfarist account of enhancement.

Chapter 7

Finally, in chapter seven, I take a global perspective on the development and distribution of enhancement and suggest a utilitarian mechanism that can mitigate inequalities arising from enhancements. Here I critique Allan Buchanan's (2011) Global Institute for Justice in Innovation, an organization he proposes as a means for speeding up the diffusion of technologies globally in order to avoid unjust distribution. I argue that it would fail to adequately face the worry that emerging enhancements may exacerbate extreme global inequality. Instead, I propose an expanded version of Aidan Hollis and Thomas Pogge's Health Impact fund (2008). The Health Impact Fund is designed to reward pharmaceutical innovators on the basis of their drugs' contribution to the lessening of the global burden of disease. Using the welfarist account of enhancement, I show how we can go about expanding the fund into a "Well-Being Impact Fund". On this scheme, innovators would be rewarded on the basis of their innovation's contribution to overall well-being – regardless of whether that innovation restores or augments some bodily or mental capacity. This quasi-utilitarian mechanism, I will argue, can respond to the concern that enhancements may exacerbate inequalities within and between societies.

To further summarize the structure of my thesis, chapter one lays out my methodological approach, chapters two to four deal with the welfarist account and its implications, chapter five begins to link it with policy matters by exploring smart-drugs, while chapter six and seven attach the account to utilitarianism to answer distributive questions raised by enhancement. Altogether, by the end of this thesis my goal is to have provided and defended an alternative account for defining enhancement that is rooted in well-being, and then using that welfarist approach to also respond to questions about how we ought to develop and distribute enhancements.

I would like to conclude with some notes on publications related to this thesis: Chapter two is largely composed of two manuscripts I published as a doctoral candidate, one in the *Journal of Medical Ethics*, and the other in the journal *Performance Enhancement and Health*:

Zohny, H. (2014a). A defence of the welfarist account of enhancement. *Performance Enhancement & Health*, 3, 123–129.

Zohny, H. (2016). Enhancement, disability and the riddle of the relevant circumstances. *Journal of Medical Ethics*, medethics-2015-103229.

Significant portions of chapter five have also been published in the journal *Neuroethics*:

Zohny, H. (2015). The myth of cognitive enhancement drugs. *Neuroethics*, 8, 257-269.

Initial thoughts that I further develop in chapter six and seven were published in a book review in the *Journal of Applied Philosophy*:

Zohny, H. (2014b). Humanity Enhanced: Genetic Choice and the Challenge for Liberal Democracies by Russell Blackford, 2013 Cambridge, MA, PB - MIT Press 248 pp., £20.95. *Journal of Applied Philosophy*, 31, 326–329.

I presented key arguments from chapter seven at the Australasian Association of Philosophy conference:

Zohny, H. (2016). *Global Justice and Enhancement: An Expanded Health-Impact Fund*. Paper presented at the Australasian Association of Philosophy Conference, Melbourne, Australia.

I also published some further ideas related to this thesis for more popular audiences in *Scientific American*:

Zohny, H. (2015, July). The Problem with Artificial Willpower. *Scientific American*. Retrieved from <https://www.scientificamerican.com/article/the-problem-with-artificial-willpower1/>

Zohny, H. (2016, May). Does Zapping Your Brain Increase Performance? *Scientific American*. Retrieved from <https://www.scientificamerican.com/article/does-zapping-your-brain-increase-performance/>

Finally, some of my key conclusions related to chapter five were also written about in *The Chronicle of Higher Education* and *The New Zealand Herald*

Bartlett, T. (2015, May). Do Smart Drugs Exist? The Answer Isn't So Simple. *The Chronicle of Higher Education*. Retrieved from <http://www.chronicle.com/article/Do-Smart-Drugs-Exist-The/229879/>

Gibb, J. (2015, May 27). "Smart" drugs not such a clever move - research. *New Zealand Herald*. Retrieved from: http://www.nzherald.co.nz/nz/news/article.cfm?c_id=1&objectid=11455399

Chapter 1

Preliminaries

In this preliminary chapter, I set out some of the caveats relevant to my approach to this topic, as well as highlight my methodological guidelines for the thesis. I will first elaborate on some basic assumptions, before exploring the topic of speculative ethics and its role in my thesis. I conclude the chapter by making a distinction between abstract and concrete ethics as a way of highlighting the methodological guidelines I intend to follow throughout the thesis.

The overarching method I use in this thesis is that of rational analytic practical ethics. Practical ethics is a linking discipline to the extent that it aims to bridge theory and practice (Thompson, 2007). As such it is characterised by drawing on a variety of disciplines in addition to philosophy, and in the context of this thesis, those other disciplines will primarily be empirical sciences. However, perhaps more central than its inter-disciplinary nature, practical ethics is also political in nature: it seeks to find workable solutions to concrete problems entangled in often messy, unpredictable and emphatically unideal circumstances. To do this successfully may often entail trying to circumvent entrenched theoretical disagreements and reach compromises that can appeal to both sides of a given divide.

To that end, one implication of my approach to this thesis is that, in using an account of a particular concept or ethical theory, I will often justify it by showing how it is compatible with other views about that concept or theory. Hence, for instance, in the context of the welfarist account of enhancement, rather than arguing for a particular theory of well-being as the correct one, I will argue that a successful account in the context of policy is likely to be one that overlaps with the core elements of most theories of well-being. This approach of seeking an overlapping consensus is key in the context of policy making: we need conceptions that are likely to be affirmed by those with opposing views (Rawls, 1987), and in that sense an overlapping consensus about a concept or idea is one that aims to capture the spirit or essence of different views about it. My appeal to utilitarianism later in the thesis is grounded in this same goal: while many abhor its theoretical implications, its recommendations in the real world and at an institutional level (as opposed to as a personal morality) tend, I argue, to be compatible with decidedly non-utilitarian views. Nevertheless, this is a caveat that is worth keeping in mind moving forward: for some non-consequentialists, this approach may seem to ignore certain

sources of value, or neglect certain ways of responding to value more generally (such as honouring rather than promoting it – more on this in chapter six).

There is another caveat to this thesis that some may find objectionable. It is that my thesis can be construed as being motivated with a speculative concern about future technologies. Yet speculation of this kind has come under criticism. In the next section I elaborate on that criticism and respond to it.

1. Enhancement and the risk of speculative ethics

If in the future we develop a technology that vastly enhances our cognitive abilities, we should ...

Attempts to complete the rest of this sentence often entail venturing into the domain of speculative ethics. *If* drugs that manipulate every mood, belief or state of mind are available in the future, will we be able to handle such a degree of self-mastery (Owens, 2007)? *If* parents can choose their children's traits, should they be held responsible for the traits their children have (Sandel, 2007)? *If* enhancements become ubiquitous, will this promote a high risk, coercive culture that reinforces negative individualist value (Meulen, 2015)?

Together, the future-based orientation of the above questions, along with the uncertainty they are based on, reflects their anticipatory nature. A premise lies behind them that sees science leaping ahead with world-transforming advancements: if the ethics doesn't keep up, we may end with a deeply troubling world (Mnyusiwalla, Daar, & Singer, 2003). And so, on a topic like the future of science and technology, we can't be too ethically prepared: the very future of humanity may depend on us getting it right from the start.

But a dilemma also lurks behind such attempts at anticipation: controlling the impact of technologies is easier when we investigate those innovations at their earlier stages, and hence why speculating about them and anticipating them can be useful. On the other hand, it is precisely at those early stages when we lack the necessary knowledge to enable us to constructively influence the course those technologies take (Collingridge, 1980). This is the so-called Collingridge Dilemma, and for those who appreciate just how much knowledge we currently lack about such prospective technologies, the whole speculative ethics project seems futile and wastes a precious resource: trained ethical deliberation (King, Whitaker, & Jones, 2011; Nordmann, 2007; Nordmann & Rip, 2009). Not only that but it may also distort our thinking – for instance, it has been argued that giving too much weight to speculative scenarios

in decision-making has led some to adopt an extreme and incoherent version of the precautionary principle (Manson, 2002).

To an extent, my thesis sidesteps the Collingridge Dilemma by not setting out to promote any specific regulations for emerging or future technologies. As stated in the introduction, my goal is to provide a framework for thinking about enhancement and its distribution; it is not to predict what specific technologies are likely or unlikely to emerge in the future. Nevertheless, while I argue for a broad understanding of enhancement that includes existing therapies and non-technology based interventions in the body and mind, there is no doubt that the thesis is also driven by a concern about what we ought to do about emerging enhancements. In that sense, I am anticipating their emergence, and secondly, presuming that they raise ethical questions or concerns. To that extent, I am arguably engaging in, if not motivated by, speculative ethics and hence it is important I respond to concerns about it.

Perhaps the most damning criticism of it comes from Alfred Nordmann (2007). He claims that the disproportionate focus of ethical concern on highly speculative scenarios reflects “what one might call the ‘if and then’ syndrome”. As he frames it:

An if-and-then statement opens by suggesting a possible technological development and continues with a consequence that demands immediate attention. What looks like an improbable, merely possible future in the first half of the sentence, appears in the second half as something inevitable. And, as the hypothetical gets displaced by a supposed actual, an imagined future overwhelms the present. (Nordmann 2007)

The worry is that speculative ethics starts off by raising legitimate questions about the future before confusing vague possibilities with impending inevitabilities that carry present ethical obligations. Nordmann (2007) gives the example of radical life extension:

The true and perfectly legitimate conditional ‘if we ever were in the position to conquer the natural ageing process and become immortal, then we would face the question whether withholding immortality is tantamount to murder’ becomes foreshortened to ‘if you call into question that biomedical research can bring about immortality within some relevant period of time, you are complicit with murder’ – no matter how remote the possibility that such research might succeed, we are morally obliged to support it.

Nordmann, in other words, fears speculative ethicists are prone to mistaking mere possibilities with imminent inevitabilities that require strong moral stances. But is that an inevitability for speculative ethicists? Surely we can raise those legitimate conditionals without necessarily deeming them as impending and in need of present action.

However, Nordmann thinks that merely raising them can be problematic: the cost of being distracted by these “incredible futures” is that it squanders the scarce resource of ethical deliberation. In other words, given the limited number of ethicists, and given the significance of current, urgent moral problems, for ethicists to focus on mere future possibilities is a poor use of their expertise. Indeed, in a world of such vast health and income inequalities, harmful exploitation, famines, as well as myriad, concrete ethical issues posed by existing technologies such as deep brain stimulation (see Gilbert & Goddard, 2014), merely spending time engaged in speculative ethics appears unethical itself.

To this, the speculatively-inclined may respond by noting that scientific research and expertise is itself scarce; speculative ethics is necessary in order to ensure we do not squander that research developing technologies that may turn out ethically unacceptable. An example of this argument has been made by Rebecca Roache (2008), who defends speculative ethics as playing an important role in helping us stop harmful future developments from unfolding. Much of our everyday ethical thinking, she notes, is rooted in anticipating and assessing possible futures, and many of those futures never unfold. Moreover, she asks us to consider examples of technological developments arising in response to ethical concerns, such as carbon capture and storage technologies responding to concerns about climate change, and advances in preimplantation genetic diagnosis technology responding to concerns about children with inherited diseases.

However, as King et al. (2011) point out, these examples relate to current, not speculative, moral problems. Anti-speculation ethicists would not, presumably, consider it unacceptable to devote our scarce moral deliberations to these issues. King et al. (2011) take this to mean, in line with Nordmann (2007), that there is an issue of distributive justice when it comes to ethical resources:

Since those engaging in speculative ethics are doing so at the expense of addressing real (i.e. not imaginary) moral issues, there is a distributive justice problem here. This allows ongoing moral problems to persist, whether these be problems related to famine, harmful exploitation of the vulnerable, or health inequalities. These wrongs and the suffering they cause are immense and are currently occurring. The obligation to use moral reflection to address these problems ought to be a concern for every moral philosopher, motivating them to seek as just a distribution of their discipline's work as possible. If these problems and many like them are taken seriously, they lead to a commitment to work on problems like these rather than on highly speculative ones.

It is worth highlighting that much of this disagreement may come down to what is considered “highly” speculative. For the moral philosopher who believes we have good reasons to expect

radical enhancements or AI to be looming and virtually inevitable, engaging with their ethics may not entail being “highly speculative” at all. But for the moral philosopher who maintains that such possibilities belong only in the realm of “incredible futures”, devoting any present thought to it may indeed be wasteful.

If this is true, then resolving the matter may largely boil down to thinking more carefully about the nature of plausibility: what makes a future technology plausible as opposed to unlikely? One challenge for supporters of the view that particular technological visions are virtually inevitable comes from science and technology studies (STS). Such views, it is argued, take technological developments for granted and fail to acknowledge the complex interplay between science and society, and thus fail to engage with the idea of science co-evolving with society in unpredictable ways (Cooper, 1999; Fox & Swazey, 2008). Acknowledging this co-evolution entails a careful analysis of myriad interacting socio-cultural, economic, political and historical factors, an analysis which often leaves us uncertain. Rejecting such uncertainties ignores the open-ended nature of the future and the possibility of alternative developments (Ferrari, Coenen, & Grunwald, 2012).

To avoid this, scholars working in STS and philosophy of technology have developed a number of approaches to evaluating emerging and future technologies that highlight the unpredictability of technological development. Examples of these include ethical technology assessment (Palm & Hansson, 2006), anticipatory technology ethics (Brey, 2012) constructive technology assessment and socio-technical scenarios (Rip & Te Kolve, 2008). All these approaches are rooted in the idea that the assessment of technological vision should explicitly acknowledge the complex and contingent nature of technological developments and their effects.

Ultimately, while I will not engage in such specific assessments, I will take a cautious approach that is not rooted in assumptions that specific future technologies are certain or largely inevitable: the strength of my case for the welfarist account of enhancement, and specifically my proposal for mitigating inequalities that may arise from enhancement, will not rely on any specific technologies emerging. In other words, my arguments will not presume any inevitability or certainty about what enhancements we will develop, or how quickly – in fact, as I demonstrate in chapter five, I critique those who do so in regards to so-called smart drugs.

This point aside, however, there is a more pertinent question: does engaging in speculative ethics actually come at the expense of addressing existing moral issues – as King et al. (2011) and Nordmann (2007) claim?

The view of moral concern as a scarce resource that requires just distribution invites an understanding of ethical concern as a pie – the bigger the slice speculative ethics takes, the less there is left for non-speculative topics. But this seems to presume that speculative ethicists are also moral philosophers or, more specifically, applied ethicists who have simply been lured away by fantastical future possibilities. Is this true?

Armin Grunwald (2010) argues that speculative ethics (particularly in regards to the domain of nanoethics, where much of this debate has unfolded) is not only conducted by moral philosophers. Instead, speculative issues have drawn in philosophers and science and technology researchers who were not otherwise engaged in applied ethics at all. That is, it is not that speculative ethics is now getting a bigger slice of the pie, but rather, the pie itself has grown by an influx of outsiders. It may be true that, as King and colleagues (2011) claim, moral philosophers *should* give greater weight to ongoing moral problems, but what King et al. do not seem to consider is that those engaging in speculative ethics are not necessarily moral philosophers to begin with, but philosophers of technology, anthropologists, sociologists and others engaging in what Grunwald (2010) suggests we call “explorative philosophy” rather than speculative ethics. While he does not provide empirical evidence for this claim, given the breadth and diversity of scholarly journals that engage with this topic, it is not an implausible assertion to make.

To the extent that this is true, moral philosophers may nevertheless complain that it is unlikely to be fruitful for outside experts to invade their turf with speculations. But what is interesting to point out is that much of speculative ethics does not necessarily entail “doing ethics” at all, but engaging in conceptual analysis¹. In asking the question of what we ought to do if a future technology enables us to vastly enhance, say, our cognition, we may find that we need to first become clearer about what we mean exactly by “enhance” and “cognition” and what those two together might entail precisely. That very process of reassessing and sharpening these concepts in the hope of better understanding the implications of possible future technologies may prove useful, not only for speculative ethicists, but for current ethical problems. In fact, rethinking concepts for anticipatory reasons can help redirect ethics to the present while offering novel ways of tackling ongoing problems.²

¹ This is not to suggest that conceptual analysis is not a crucial part of what moral philosophers do. But the task of conceptual analysis in general is clearly not one exclusive to moral philosophers.

² A great example of this is speculating on the ethics of generation starships (Levy, 2016). These are spaceships that set out to colonize other planets, but due to the distance of travel, they must support not just those who set out on them, but also their descendants. Is it permissible to give birth to children within such a constrained domain, where life projects are limited and one is bound to the ship, and where one

In summary, to the extent that the concerns about speculative ethics are persuasive, my arguments will not rely on any claims of inevitable or impending innovations. On the other hand, much of my thesis, like much of the literature on enhancement, is engaged in conceptual analysis, setting the ground for applied ethics, as opposed to directly engaging in applied ethics – something that arguments against speculative ethics do not seem to take into account.

However, while limiting the role that speculation will play in my thesis is one methodological guideline I will follow, there are others. In the next and final section of this chapter, I elaborate on those.

2. Methodological guidelines

Here, I want to make a case for a reliance on empirical evidence in reaching conclusions throughout this thesis, as well as a case for distinguishing between abstract and concrete ethics where possible. First, let us distinguish abstract from concrete ethical questions. Consider the difference between these two questions (Maslen, 2015):

1. Suppose a safe drug X is developed that is highly effective at keeping users vigilant and fatigue-free. Should surgeons take it?
2. Should the American College of Surgeons recommend the use of modafinil for its members?

The first question is interesting: answering it requires us to answer questions about the limits of professional obligations, the relationship between cognition and responsibility, and specifically how the burden of responsibility may increase as we gain more control over our levels of fatigue and so on. More broadly, it invites us to analyze what we mean by cognition and responsibility. Answering this question is undoubtedly very useful – we may need to answer it if we can ever hope to answer the second question (Maslen, 2015).

And yet, no matter how full an answer we develop for the first question, it will not be sufficient for answering the second question. This is because “safe drug X” is different from modafinil. Modafinil may turn out to be an all things considered “safe” drug that is effective at keeping users vigilant, but we also know it can cause anxiety, gastro-intestinal upset, and can leave users

is pressured if not forced to reproduce? Generation starships are not impending possibilities, but answering the questions they raise can freshen our perspectives on what we owe people on planet earth, where many children are born locked into poverty and in a world with a growing environment crisis.

feeling over-confident (Baranski, Pigeau, Dinich, & Jacobs, 2004a; Teitelman, 2001). This means that we cannot expect the answer to the abstract, detail-free scenario in the first question to hold up when we consider the details entailed by the concrete scenario in the second question. The reality of these side-effects means that we now have the added, and in a sense, more complicated, task of weighing up indeterminate risks (how likely and to what degree might a particular surgeon feel over-confident if they use this drug?) with indeterminate benefits (is a surgeon actually less likely to make a serious mistake while on modafinil)? Such a more complicated, concrete scenario will bear differently on the limits of professional obligations, on how effective we think the drug is, and on whether it is worth the risks. We cannot expect the answer to the first question to hold up as an answer to the second one.

In other words, while answering the first question can be hugely useful for thinking clearly about the core, abstract issues, it will not, in and of itself, provide us with an answer to the second question. As I will show in chapter five, question two becomes even more complicated when we simply do not know what the efficacy and side-effects of drugs like modafinil are. And yet, we have far more information about the risks and benefits of modafinil than, say, implanting computer chips into our brains. So, while a version of question one for implanting computer chips may still be useful (is it, in principle, permissible and what concepts do we need to reach such a conclusion?), answering it will tell us even less about whether we should be attempting that – we would need far more details about their risks and benefits first.

It is worth pointing out that we could arguably reclassify these two questions as ones that respectively pertain to theoretical and applied ethics more generally. However, to reduce potential confusions about how these relate to practical ethics, which as I state is my overarching methodological approach to this thesis, let me be clear: while practical ethics is ultimately concerned with resolving questions of the concrete or applied variety, I take it for granted that to do so effectively requires also having coherent answers to abstract ethical questions. Hence, in doing practical ethics, I will be engaged in both types of questions. However, what is crucial here is that we should be clear about which type of ethical question we are tackling in any given discussion: abstract or concrete. Not doing so can lead to needless ambiguities and disagreements – something I highlight especially in chapters four, five and six. As such, one of my methodological goals for this thesis is to make it clear which of those types of questions I am engaging in.

There is a further implication of this distinction between abstract and concrete ethical questions: when bioethicists are addressing questions of the second type, this should be informed by

assessment and validation of empirical findings. In the case of a drug like modafinil, this would entail an ability to evaluate the state of the evidence for its efficacy, safety and prevalence. Indeed, in chapter five I argue that a lack of rigorous engagement with empirical data on drugs like modafinil has led to a degree of hype in the cognitive enhancement literature that threatens to undermine the credibility of bioethics, and specifically its ability to accurately assess the ethical dimensions of emerging technologies.

Further linked to this is a failure to acknowledge underlying assumptions in the human enhancement literature. These assumptions often have a determining impact on the normative stances taken by scholars writing on the topic (Ferrari et al., 2012). Often they relate to implicit empirical claims – such as claims about the efficacy or prevalence of, say, putative cognitive enhancers – which give a false sense of urgency to these debates. Sometimes, they are normative assumptions, such as whether one is thinking in purely consequentialist terms or not. As I have stated above, one assumption I am making throughout this thesis is that I consider thinking in terms of expected harms and benefits to be the relevant way to think about the issues raised here.

Altogether, in light of these issues, I strive to ground my research in as much empirical data as possible, and to state clearly any assumptions I make. For instance, in the next two chapters, I limit my discussion largely to conceptual analyses. To the extent that I consider ethical issues, they remain abstract and only serve to merely aid the process of defending and refining my approach to the welfarist account of enhancement. However, in chapter four and five, I start to increasingly rely on empirical evidence as I investigate the relationship between moral behavior and cognitive ability, and subsequently the evidence for so-called smart drugs such as modafinil and methylphenidate. In chapter six, in arguing for a utilitarian approach to the distribution of emerging enhancements, I again revert to largely abstract ethical questions, before concluding the thesis with a concrete proposal rooted in the ethical and conceptual analyses carried out in the preceding chapters.

To summarize, I root my methodological approach to this thesis in these ideals:

- 1) Avoiding highly speculative scenarios rooted in claims of inevitable or impending technologies.
- 2) Being clear on when I intend to engage in abstract as opposed to concrete ethics.
- 3) Using of ethically relevant empirical findings to inform my reasoning where possible.
- 4) Acknowledging assumptions or hidden premises clearly.

With that in mind, let us proceed with an investigation of that perennial question: what is enhancement?

Chapter 2

Enhancement and Well-Being

Imagine taking a performance-enhancing drug that allows you to solve math problems a lot faster than you used to¹. If your math skills were previously average, we could say that the drug has enhanced you in that respect. If your math skills were previously very poor (say, due to some identifiable cognitive dysfunction), and now you are able to solve maths problems at an average speed, we could say the drug has treated you. Based on this distinction, an enhancement seems to be an intervention that goes beyond the ends of medicine in what is necessary to sustain or restore good health or normal functioning. It's a common definition of enhancement (Daniels, 2000; Juengst, 1998; Pellegrino, 2004; Resnik, 2000). We may call this the *transmedical approach*.

Alternatively, we can ignore whether your math skills are average or poor and just focus on the fact that the drug boosts your skills compared to how they were before. In that sense, enhancement can be understood in terms of its augmentative impact on your capacities, regardless of whether they are below or above a certain standard of functioning relative to a given norm or population (Bostrom & Roache, 2008; Engelbart, 1962; John Harris, 2007). We may call this the *augmentative approach*. Together with the transmedical approach, we can think of these accounts as *functionalist* approaches of enhancement: they see enhancement as ultimately about improving our functioning.

There are other possible ways of defining enhancement, some of which have been called *ideological* – these are characterized by applying a controversial set of values to a somewhat arbitrary list of technologies that are deemed relevant to enhancement – and others that can be descried as *social constructionist* – where what constitutes an enhancement is entirely determined by what a culture values in terms of performance (see Camporesi, 2014; Savulescu et al., 2011).²

¹ A modified version of my writing from here until section three of this chapter has been published while working on my thesis: Zohny, H. (2014b). A defence of the welfarist account of enhancement. *Performance Enhancement & Health*, 3(3), 123-129. Any modifications I have made have been to make the introduction and conclusion cohere with the broader aims of the thesis.

² Outlining the different accounts of enhancement is well-trodden terrain, and so I spend little time doing so here. Instead, I engage with these different accounts only as they become relevant to my arguments for the welfarist account.

However, I am interested in a particular account that understands the term to refer to a change in the body or mind that tends to improve one's well-being. This is the welfarist account of enhancement, largely developed by Julian Savulescu and colleagues (Earp, Sandberg, Kahane, & Savulescu, 2014; Savulescu, 2006; Savulescu et al., 2011). It defines an enhancement as:

Any change in the biology or psychology of a person which increases the chances of leading a good life in the relevant set of circumstances. (Savulescu et al., 2011)

Based on this definition, the performance-enhancing drug envisioned above is not necessarily enhancing at all. This is because, whether it qualifies as enhancing or not depends on whether having better math skills would likely contribute to your well-being – that is, to your chances of leading a good life in your set of circumstances, which I understand to refer to one's particular societal and personal context (although more on this in section three of this chapter).

As I noted in the introduction to the thesis, I believe this account can provide a framework for thinking more clearly about enhancement, using the value of well-being as a common currency that bio-conservatives and their opponents can use to develop, articulate and defend their arguments. While I presented some preliminary rationales for the account, I develop these more fully here. I do so over the next two sections, although the full implications of the account will become clearer as I present and respond to objections of the account in section three. In section four, I explore the implications of “the relevant set of circumstances” in the definition by looking at how this account ties in with the concept of disability and the interplay between social forces and well-being. First, a brief methodological note linked to the conclusion of the previous chapter: this chapter is purely concerned with a conceptual analysis of enhancement. To the extent that I consider ethical issues, they are abstract in nature. My consideration of them is merely a means for aiding the process of formulating a coherent concept of enhancement.

1. The welfarist account and its critics

The welfarist account of enhancement is, to an extent, motivated by seeking an overlapping consensus about the concept. Dictionary definitions of “enhance” typically include “to raise or increase in price, value, importance;” or an “improvement in quality, value, or extent” (Simpson, & Weiner, 1989). As Savulescu (2006) points out, the spirit of such definitions and others is to *increase in value*, which, in the context of human lives, refers to the value of an individual's life for that individual. The welfarist account, in that sense, considers enhancements as those changes that increase the value of a person's life.

Those changes could be anything: the use of “any change in the biology or psychology of a person” entails that the account disregards the distinction between what we might think of as science or technology-based interventions in the body or mind (such as taking a drug or inserting a brain implant), and more conventional interventions (such as exercise or education).³ “Any change” also entails that whether the intervention is for treating a dysfunction or for augmenting an already well-functioning capacity, both can be considered enhancements – so long as they tend to increase well-being. As such, therapeutic treatments are rendered a subclass of more general enhancing interventions.

However, because of this, the welfarist account has been criticized on multiple grounds. Sparrow (2013) argues that these implications mean that the term becomes very broad, losing contact with the everyday language use of enhancement. Coenen, Schuijff, & Smits (2011) similarly suggest that enhancement may become a useless term if it fails to highlight science-based interventions and distinguish them from more conventional ones. They also argue that having a concept of enhancement that allows us to clearly distinguish it from therapeutic interventions will be key to effective policy-making.

Other criticisms have centred on its reliance on the contested notion of “a good life” and well-being more generally. Beck and Stroop (2015) argue that this account will not be helpful in classifying interventions as enhancing or not due to the variety of conceptions of well-being. There is also a worry that, by defining enhancement as whatever will likely contribute to well-being, this can act like a philosophical sleight of hand that appears to sidestep the ethics of enhancement: if to enhance is to increase well-being, then it may seem difficult to conceive of anything objectionable about enhancement (Beck & Stroop, 2015; Sparrow, 2013). Further criticisms claim that this account risks underestimating the social and cultural pressures that influence individual choices, as well as fails to take into account the social and collective consequences of enhancing (Camporesi, 2014; Gordon, 2014).

These criticisms are considerable. However, I argue here that they are either unfounded or addressable. To see how, we first need to more fully engage with the implications of the welfarist account.

³ I use changes in the biology and psychology of a person interchangeably with bodily and mental changes.

2. A deeper understanding of the welfarist account

A helpful way of understanding the welfarist account is to start by making a distinction between functionalist enhancement and what we can call welfarist enhancement (Savulescu et al., 2011)⁴. As noted, enhancement in the functionalist sense denotes some improvement in a capacity, regardless of whether we are using a transmedical approach or augmentative one. This says nothing about the well-being impact that improvement has on the person undergoing it. In general, we tend to associate improved capacities with positive effects on our quality of life. Better memory, more resilient immune systems, stronger bodies – these can implicitly suggest a corresponding improvement in the quality of one's life.

But we can also see how that won't always be the case – there is no necessary connection between these capacities and well-being. If having improved memory led to a greater propensity to developing phobias, or if improved hearing led to being constantly distracted, these may well result in a diminishment to one's quality of life. Similarly, we can see how diminishing our capacities to recall certain traumatic events, or even our ability to hear if we lived in an incredibly noisy environment – these might lead to an increase in one's quality of life, even though they would be the opposite of functionalist enhancement (Earp et al., 2014). More starkly, none of the alterations to these cognitive or sensory systems would be relevant to quality of life if the person undergoing them is in a permanently vegetative state.

It is this potential discrepancy between functionalist enhancement and well-being that, in part at least, prompts the notion of welfarist enhancement. Here, it is not the change to our capacities that constitutes the enhancement, it is the positive impact of that change on our well-being. Hence, if super hearing abilities were to lead a person to constant distraction, thereby reducing their well-being, that would not constitute an instance of welfarist enhancement. Similarly, if someone suffering from PTSD *diminished* their ability to recollect certain events and thereby increased their well-being, that would be an instance of welfarist enhancement (Earp et al., 2014).

This is why the welfarist definition of enhancement emphasises *any change* in the biology or psychology of a person: it does not have to be a strictly functionalist improvement, but may in

⁴ Savulescu, Sandberg and Kahane (2011) use the term “human” enhancement rather than welfarist enhancement. However, the use of “human” is a misnomer: depending on how we define personhood, some humans may not count as people and are not capable of well-being, and arguably some animals may count as people, or at least as candidates for welfarist enhancement, to the extent that they have a well-being.

fact entail a certain diminishment of function. For the welfarist account, what matters is the impact of the functional improvement or diminishment on a person's chances of leading a good life.

This also relates to the account's emphasis on the "relevant set of circumstances" of the person undergoing an intervention – which is to say, the person's particular context. What this entails is that there is no context-independent answer to the question of whether an intervention is an enhancement. The circumstances of the individual undergoing an intervention play a determining role in whether that intervention will likely be enhancing. In the case of the drug that boosts maths skills envisioned above, if one works as an accountant, it *may* be enhancing for them, even if it had the side-effect of, say, reducing their musical abilities. On the other hand, if a musician who rarely works with numbers used the drug, it may be difficult – in general – to see how the drug is enhancing, given their particular circumstances. I return to further implications raised by this emphasis on the relevant circumstances in section three

Note that these elaborations highlight how this is an account that defines what an enhancement *is* – it is not an account of whether that enhancement should be made use of. That is, the fact that a biological or psychological change may be enhancing in the sense of increasing well-being, is not to say that it should be undergone, all things considered. Depending on the alternatives, it may be that changing the environment instead of the individuals' biology or psychology will be more reasonable or practical. Moreover, it may be that making use of an enhancement will result in harm to others or it may create or exacerbate injustice more generally. The welfarist account is compatible with these possibilities: the fact that a modification is enhancing does not necessarily mean that it should be made use of.

As we will soon see, many of the objections to the welfarist account noted above lose their force once we have properly understood the implications of this approach. Before moving onto those, it is worth noting that while I have highlighted the preliminary implications of the role played by "any changes" and the "relevant set of circumstances" in the definition, I have not mentioned what is meant by a good life or an increase the chances of having one. While I return to the topic of well-being in later sections, it is worth adding a few remarks on the topic here before proceeding.

I take leading a good life to be synonymous with leading a life of well-being – the more well-being, the better the life. However, the term "good life" can be misleading, as it can be interpreted as "a morally good life" or at least a life of well-being that is also moral. However, for the purposes of this thesis, and in line with most theories of well-being, I limit the meaning

of a good life to solely what is good *for* a person having that life. This variety of good is called prudential value (Crisp, 2016), which we can contrast with moral value – things that are good from the view point of morality, as opposed to a self-regarding view point. In that way, the welfarist account is rooted in prudential value: it is about what is good *for* a person in terms of changes to their body or mind. This goodness may be expressed in terms of pleasure, meaningfulness, the satisfaction or fulfilment of their desires, the attainment of other goods like knowledge or friendship – but it does not include what may be good independently of that person’s well-being. In that way, it is not a moralized account of enhancement – that is, it is not concerned with moral value as contrasted with prudential value.

With that in mind, it should now be clearer what increasing the chances of leading a good life means: it means that one is more likely to have the kind of life where things that are good for them take place. Of course, how we assess such an increase in chances is a further matter, as is whether we should evaluate such an increase across a person’s entire lifespan or in a given moment or specific period. These further questions would, presumably, be answered by the theory of well-being we adopt. For our purposes, we need not worry too much about these issues at this time (I return to questions of well-being in chapters six and seven).

3. Objections to the welfarist account

3.1 Renders “enhancement” too broadly

This criticism notes that, by focusing on *any* change in a person’s biology or psychology, the welfarist definition doesn’t distinguish between science or technology-based enhancements into the body – such as drugs or genetic manipulations – and more conventional enhancements – such as exercise or education (Blackford, 2012). Coenen and colleagues (2011) argue that there are qualitative differences between the means used in science-based and conventional enhancements, and that these should not be defined away. While they don’t clarify what those qualitative differences are precisely, Wagner and colleagues (2015) elaborate on them by contrasting cognitive enhancement drugs with education.

They acknowledge that, like drugs, education impacts synaptic connections and neural networks in the brain. However, they argue that education does so in a way that is more stable and long lasting. This is because, unlike the effect of drugs, neural changes due to education are the result of repetition and practice over an extended time frame. They argue that a definition of

enhancement that ignores such differences risks committing a fallacy of ambiguity (Wagner et al., 2015).

While Wagner and colleagues (2015) are focusing specifically on pharmacological cognitive enhancement, their argument relates to a broader intuition that science-based enhancements, like drugs, work in a way that is transient and effortless, while conventional enhancements like education and training require effort and have more stable and long lasting effects. However, this dichotomy between transient and effortless versus enduring and effortful does not hold up.

Let's first consider the case of transience. On the one hand, it's certainly conceivable that future cognitive enhancing drugs will have longer lasting or potentially permanent effects. For instance, a single high dose of the hallucinogen psilocybin (the active ingredient in magic mushrooms) can create lasting personality changes (MacLean, Johnson, & Griffiths, 2011). Beyond drugs, the effects of genetic modifications would be anything but transitory. Similarly, bodily or neural prosthetics entail a degree of permanence. In contrast, despite the effort and repetition involved in studying, few go on to remember the details of a class mere weeks after exams. And despite the effort involved in endurance training, fitness levels start to plummet in as little as two weeks of not training (Coyle, Hemmert, & Coggan, 1986). Which is to say, conventionally gained knowledge and fitness can be transient too.

What about the claim that science-based enhancements entail less effort? Again, we can consider a number of counter-examples. Firstly, taking a putative cognitive enhancing drug will not, in and of itself, lead to performing better in, say, an exam. Study is still required: we still have to learn the relevant information. Similarly, taking steroids has to be combined with training to build muscles, and individuals fitted with prosthetics or a brain-machine interface need weeks of practice before they can use them properly. In contrast, it's not always clear that conventional enhancement involves much effort: watching entertaining documentaries can be highly educational, yet it barely requires effort (unless we are talking about a more superficial kind of effort, such as dedicating time to watching, in which case we can also talk about the effort of remembering to take a pill every day or the effort of hooking up a brain-machine interface).

More generally, it is not clear what makes an intervention science or technology-based to begin with. Modern education and training have become deeply intertwined with technology, and so has how we produce our food – good nutrition having obvious enhancing effects on our bodies and minds.

Nevertheless, it is true that things like transience and effort can be morally relevant in this context. Investing in a life that is built around transient effects along with a diminished sense of accomplishment may be a less valuable sort of life.⁵ But while we may be disposed to thinking that science-based interventions raise more pressing questions about health safety, authenticity, accomplishment, societal values, distributive justice, and the host of other ethical issues associated with enhancement technologies, these questions are also relevant to so-called conventional enhancements. For instance, we can ask questions about the health safety of intensive study, the stress and sedentary lifestyle it entails being issues of growing health concern. Also, education can change people in profound, irreversible and often rapid ways, and hence it may also raise questions related to authenticity and societal values. In fact, reading a single book can have profound and irreversible effects on a person. More broadly, how we regulate education, access to nutrition, and healthcare – all things that can have life-changing enhancing effects on individuals – has significant implications for fairness and distributive justice.

And so, the fact that these issues – safety, authenticity, societal values, and justice – may appear to arise in seemingly novel or pressing ways when thinking about current or hypothetical interventions associated with science-based enhancements does not make them *qualitatively* different issues. The intuition that they are is a misplaced one, and the welfarist definition helps us sidestep that false intuition. In that sense, the account keeps the focus on what is actually relevant: the impact of interventions on our lives rather than the means behind those interventions.

The implications of this for the thesis is to reassert the broadness of its scope: enhancement is a larger class of interventions than it is ordinarily thought to be. In saying that, we may have other reasons to limit our focus on particular types of enhancements that fall under some specific methods for enhancement, such as biomedical ones, for instance. However, our reasons for doing so will not be because such a method is qualitatively different from others in some fundamental way. Instead, it may be because compartmentalizing enhancements into different categories is more practical for regulating them in the context of policy-making. The welfarist account is perfectly compatible with such an end: if it is helpful for policy-makers, we can home

⁵ Although, note that if investing in a life built around transient effects and a diminished sense of accomplishment reduces the chances of leading a good life, interventions that lead to such a diminishment would not be considered enhancements based on the welfarist account.

the welfarist account onto, say, biomedical changes to the body or mind of a person, or whatever other methods of change that may be relevant to regulatory authorities.

3.2 Fails to distinguish therapy from enhancement

A similar criticism focuses on the welfarist definition's lack of a distinction between therapy and enhancement. I have engaged with this issue in the introduction to the thesis, arguing that it is not clear this distinction is tenable, and that even if it is, it's not clear what it would serve. Nevertheless, it is worth examining why some have argued for maintaining it as an objection to the welfarist account.

Coenen and colleagues (2011) describe the importance of that distinction as the key conceptual challenge in policy-making contexts, allowing us, for instance, to prioritize the relevant sorts of interventions in terms of what we research and fund. Given limited resources, most would be inclined to argue that therapies should be prioritized over transmedical enhancements since being diseased or injured is more likely to stop an individual from being a normal competitor able to lead a good life. Prioritizing therapies is therefore central to maintaining basic standards of equal opportunity (Daniels, 2000).

The welfarist definition of enhancement appears to eschew this challenge, demoting therapeutic treatments of disease and injury to a subclass of other interventions that tend to increase well-being in a given context. Hence, whether one is mending a broken arm, or cutting off a healthy arm and replacing it with a stronger, bionic one, both are enhancements so long as they tend to increase well-being in the relevant circumstances.

Adding to that line of thought, some have argued that if we do not make a distinction between therapy and enhancement, it may be too easy to argue that all forms of enhancement are morally permissible given that the things we count as therapies are permissible (Lin & Allhoff, 2008). However, it's not clear that just because something counts as a therapy it is deemed permissible. Its permissibility is contingent on the circumstances of the person undergoing it. For instance, it would not be permissible to administer chemotherapy to someone who doesn't have cancer. Therapeutic treatments, especially risky ones with adverse side effects, are generally only permissible when we have reason to believe that they will make the life of the person undergoing it go better overall. That, in the broadest terms, is our guiding criterion for such decisions.

And this is where the welfarist account is most useful. Because, while it may not provide therapy-enhancement distinction, it provides a more relevant distinction: interventions to the body and/or mind that tend to contribute to the well-being of a person, and those that do not. As it happens, things that we think of as therapies tend to contribute more to our well-being than interventions that might improve our functioning beyond some norm. That is, in most circumstances, mending a broken limb or treating cancer contributes far more to our chances of leading a good life than, say, augmenting normal eye sight or concentration levels. Under the welfarist account, this gives us concrete, normative reasons to prioritize such therapies over such transmedical enhancements. However, this will not always be the case. Some therapies will have little impact on the quality of our lives (say, a treatment for mild eczema), while some transmedical enhancements may have deeply positive impacts (say, improving one's self-control Moffitt et al., 2011). The welfarist account encourages us to acknowledge that, in some cases, developing and funding interventions that do not technically restore health may be more valuable than interventions that do. And that is because of their impact on well-being.

To that extent, the welfarist definition again keeps us focused on what actually matters: not on nominal and hazy distinctions between interventions, but on how those interventions impact well-being. However, in highlighting the centrality of well-being, this brings us to a more fundamental objection to this account – isn't how an intervention impacts well-being itself hazy?

3.3 Overly dependent on the contested concept of well-being

This criticism is well articulated by Beck and Stroop (2015):

This account, however, seems too widely based to provide a clear definition, for whether or not something counts as welfarist enhancement under this description depends on how we understand the notion of a "good life". Alternative theories of well-being and the specific aspects on which they lay emphasis are, however, likely to result in divergent classifications.

In other words, since there is no consensus about what constitutes leading a good life, there can be no consensus about what sorts of interventions are enhancing. Savulescu and colleagues (2011) acknowledge that whether an intervention can be understood as enhancing will depend on the account of well-being we employ. But is the existence of competing accounts of well-being as problematic as Beck and Stoop (2015) suggest?

First, it is worth highlighting that some interventions are clearly enhancing on any plausible conception of a good life for most people in ordinary circumstances. This is the case for the majority of interventions that aim to restore health or prevent disease. In practice, we will likely be able to identify a wide range of interventions as either enhancing or harmful to an individual without much controversy.

There is a reason for this: while the main theories of well-being—hedonism, desire-satisfaction, and objective list theories—offer conflicting accounts of the nature of well-being, they nevertheless converge to a large extent on the things that can make a life go better or worse. In fact, at least for hedonism and desire-satisfaction, their aim *is* to converge. Consider things like accomplishment, friendship, autonomy, and knowledge. Hedonism and desire-satisfaction theories typically hold these as instrumentally good because, for most people, they give us pleasure or satisfy our desires. Objective list theories in turn tend to recognize the intrinsic value of pleasure and typically hold that informed desires often reflect what is independently good (Kahane & Savulescu, 2009). To that extent, seeking an overlapping consensus about what well-being entails is an increasingly popular approach (Scott, 2012).⁶

Such an approach does not deny there will be difficult cases related to whether an intervention will be enhancing for a person. However, there is no reason to believe that these will be more difficult than cases where it is unclear whether a treatment for an illness will pay off for an individual. Health and disease are indeterminate concepts – in fact, there are no less than eight different disease concepts held by individuals engaging with clinical reasoning and practice (Freitas, 1999) – yet that generally does not pose a significant problem for determining who to treat and for what.

However, there is a problem related to the objection raised here that has not received much attention. Even if we were to form a consensus on the right account of well-being, there is still the problem of how we measure the well-being of individuals. To be sure, this is a problem for any welfarist account, not just in regards to enhancement. The problem is this: we know that individuals are fallible. People can simply be mistaken about what is good for them (Levy, 2012). Related to this problem is that, to the extent that well-being entails a subjective component based on a person's evaluation of the goodness of their life, there appears to be two ways to

⁶ Perfectionism about well-being – broadly, the idea that exercising and developing human-specific capacities is what is good for humans (Fletcher, 2016) – is often left out from these discussions of convergence. However, we can see how, to the extent that forming friendships, attaining knowledge and being accomplished reflect capacities in our nature, perfectionism would similarly converge with other theories of well-being.

measure that evaluation. We can speak broadly of life satisfaction, which implies reflecting back at an experience or series of experiences related to one's life, and emotional well-being, which relates to how we feel during an experience as it unfolds. These two modes of evaluating our well-being can diverge quite significantly, so much so that we can speak of individuals as having a "remembering self" and an "experiencing self" (Kahneman, 2011).

To illustrate this, consider the relationship between income and life satisfaction, and income and emotional well-being. While life satisfaction continues to increase as one's income increases, emotional well-being does not. According to one study in the US, unlike life satisfaction, emotional well-being stops correlating with annual income once it exceeds \$75,000 (Kahneman & Deaton, 2010). Another illustrative study compared women in Columbus, Ohio, with those in Rennes, France, and found that even though American women reported higher levels of life satisfaction, French women reported spending more time in positive moods and doing the sorts of things that both French and American women agreed yielded more pleasure (Krueger et al., 2008). Whose life should we say is going better?

This discrepancy between the evaluations of our life satisfaction compared to our day to day emotional well-being may pose a problem for the welfarist account of enhancement. It raises the question of whose well-being enhancements are meant to be improving: the remembering self or the experiencing self? It is the remembering self that ultimately makes decisions. But the remembering self is not a very good judge of what contributes to well-being and what doesn't – especially when it comes to unpleasant experiences. A study illustrating this phenomenon had participants immersing one of their hands for one minute in water cooled at 14C – which is considered painful. After, they immersed their other hand for 1 minute in water at 14C, followed by a further 30 seconds during which the water was gradually warmed to 15C. Despite the added 30 seconds of still painfully cold water, participants were more willing to repeat the second experience (Kahneman, Fredrickson, Schreiber, & Redelmeier, 1993). This is because the last 30 seconds of slightly warmer water coloured their perception of the whole experience.

This, along with similar studies (Redelmeier & Kahneman, 1996; Schreiber & Kahneman, 2000), reflects our bias towards giving too much weight to the last moments of an experience. In the context of welfarist enhancement, this raises concerns not only about our ability to decide what is actually best for us, but also in what ways we are aiming to improve the quality of our lives exactly – in ways that are primarily experienced, or remembered?

Presumably, the account of well-being that we adopt should be able to answer this question, and it's likely to be one that can accommodate both of these selves and their sometimes

conflicting interests. In reality, though, it's likely that most things that contribute to our emotional well-being will also contribute to our life satisfaction, and vice versa. It's also worth pointing out that our fallibility about our own well-being is not as troubling for this account as it might seem. It may be tempting to think that such fallibility opens the door for paternalism or coercion about enhancement (or any other issue, for that matter). However, by rejecting the idea that any person is infallible about well-being, we thereby also reject the idea that a would-be paternalist is infallible on the matter as well (Moore, 1994). In that sense, acknowledging our fallibilism can make paternalism on these matters harder to justify.

More importantly, a welfarist account that is sensitive to our fallibility is one that, to an extent, is less likely to miss important elements of our well-being. Acknowledging our fallibility and becoming aware of the conflicting ways in which we can be wrong about what is good for us should be a part of the process of deliberating about which interventions tend to contribute to leading good lives. When a welfarist account takes these issues into consideration, it is more useful than approaches to enhancement that disregard well-being altogether. Ultimately, the problem of cognitive biases and flaws in our reasoning about what will improve our well-being is an opportunity for this account. If we develop interventions that can reduce these cognitive errors, then that would be a crucial step towards increasing the chances of leading a good life in a very broad way.

3.4 Does not offer timely evaluations

This criticism has also been articulated by Beck and Stroop (2015) as:

When following the welfarist account, enhancement procedures can only be defined as such after the impact of an intervention has become apparent and has been shown to be a contributive factor to a person's well-being.

The worry here is that this approach will not be useful in regards to helping us evaluate, in a timely, pragmatic fashion, whether an intervention is actually an enhancement. It seems to imply that, for instance, in attempting to evaluate whether the off-label use of some pharmaceuticals are instances of enhancement, we can only wait and see what impact they have on wellbeing – by which time the practice may have become common place.

Yet isn't this precisely how medicine works? Medicine involves the invention of new interventions that are hypothesized to be therapeutic in their effects. Yet we do not have a problem waiting to see whether that proposed intervention has actually had a therapeutic effect

on relevant subjects in a trial setting before we deem it to be a therapeutic intervention. Even then, if the trial results are promising and we begin to call the intervention “therapeutic”, it doesn’t follow that it will actually increase health for anyone who undergoes it – but only for individuals in a relevant set of circumstances (that is, who have the condition that this intervention treats). Even in that case, we cannot be sure that it will work for that individual. Yet this uncertainty in medicine, and the retrospective element required for determining whether an intervention is therapeutic for an individual, is not seen as problematic. We can agree that an intervention will likely be therapeutic in a given context, and we can regulate it on that basis, but we accept that whether it will in fact be therapeutic for an individual is an open question.

The welfarist approach to enhancement can work the same way. A new intervention can be classified as enhancing based on studies reporting that it tends to increase broadly agreed upon notions of wellbeing for individuals in a given context, but that whether it will actually be enhancing for every individual will depend on a host of factors. The fact that some pharmaceuticals are used off-label for putatively enhancing purposes is not a problem for the welfarist account – it is a problem for regulations stemming from, in part, the lack of studies investigating whether their use is conducive to wellbeing. I return to this issue at the end of chapter five and examine how the welfarist account can interact with such regulatory issues.

3.5 Sidesteps the ethical debate on enhancement

Sparrow argues that the welfarist account risks settling the complex ethical issues of enhancement through some philosophical sleight-of-hand (Sparrow, 2013). If enhancement is not qualitatively different from therapies or more conventional enhancements like exercise, what is all the ethical fuss about? Beck and Stroop (2015) voice this worry as:

It is questionable whether the wide understanding of enhancement, which the welfarist account offers, properly meets the moral concerns critics of enhancement usually articulate. If everything that actually contributes to (the chances of) leading a good life is considered as enhancement, then it is not obvious how enhancement could be in the least objectionable.

Similarly, Camporesi (2014) argues that the welfarist definition’s focus on individual well-being neglects the broader societal and collective consequences of enhancement.

However, these concerns appear to mistakenly assume that by singling out the value of well-being at the individual level, the welfarist account does so in an exclusive way – as though other values, such as justice, are negligible. But this is false.

The welfarist approach is not meant to resolve the question of whether an enhancement is ethically permissible, it is merely meant to reframe our understanding of the concept of enhancement in a way that ties it to the well-being of the individual undergoing a modification. Well-being is a prudential good, and in that sense this account aims to capture why an enhancement is *good for* the person enhanced. But by singling out well-being in its definition, it does not follow that it is the only relevant thing when it comes to thinking about the permissibility of a given enhancement.

As Savulescu and colleagues (2011) state:

The welfarist approach distinguishes ways in which some treatment might benefit a person from other relevant values, such as justice. It thus allows us to say that although some treatment is an enhancement (i.e. contributes to an individual's well-being), it might nevertheless be bad overall, because its employment in the current social context will lead to far greater injustice.

In that sense, considerations about other people's well-being, societal impacts, distributive justice, and whatever else a moral theory might be concerned with aside from prudential good, are still relevant considerations. Highlighting individual well-being keeps the definition focused away from the terminological controversies and conceptual confusions of the other accounts. It instead keeps the debate focused on what is most relevant in making sense of the relationship between an intervention and a person undergoing it: its impact on that person's well-being.

However, it is understandable where this criticism arises from. Indeed, the crux of the welfarist approach is that it transforms the question of when to enhance to when we should increase well-being. Met with such a question, we may want to immediately answer: whenever we can. This indeed appears to be Savulescu's (2006) response to that question:

When enhancement is understood as an intervention which increases the chances of a person having a good life, it is hard to see how there could be any objections to trying to make people's lives go better. Indeed, the fact that enhancements increase well-being provides a strong moral obligation based on beneficence to provide them.

If individual well-being is independent of any reliance on shared resources, then Savulescu may well be correct here. But well-being, however conceived, is arguably dependent to some extent on access to various external goods: at the least, shelter, healthcare, and so on. It moreover depends on what one is already used to: shelter in the form of a cave for most people living in

the 21st century would be significantly diminishing to their well-being. Given this, well-being seems to share a close relationship with limited goods. How much of those goods one appropriates has implications for how much is left for others, and in that sense, the level of well-being of an individual has implications for the well-being of others. This brings up issues of distributive justice. How should access to the sorts of goods conducive to well-being be distributed? The answer to this question depends on the account of justice employed, which is the topic of the last two chapters in the thesis. Regardless, at this stage in our ethical deliberations, things may get more complicated. But at least with the welfarist account, they get complicated in a relevant way – that is, by keeping us focused on what is good for individuals.

3.6 Underestimates social forces

This final objection raised in the literature relates to the criticism by Camporesi (2014) that the welfarist account appears to underestimate the social and cultural pressures that influence individual choice and how we define well-being. For instance, in critiquing the welfarist account, Gordon argues that merely asking whether an intervention will increase well-being does not go far enough to capture all that is relevant in thinking about an intervention (Gordon, 2014). Instead, he argues we should also ask what social conditions have led us to consider a particular intervention as an enhancement – as something that will tend to increase the chances of leading a good life? And, secondly, what social structures are enabled by the adoption of a given enhancement?

These are important questions, but while they are certainly relevant to an assessment of the permissibility of an enhancement, it's not clear that they are relevant to how we should *define* enhancement. To explain why, let us assume that the social conditions that are moulding how we define well-being are nefarious ones. For instance, we can imagine a hyper-capitalist society where the basic goods that tend to be conducive to well-being can only be attained by becoming a highly productive contributor to the economy. Here, employees operate under social forces that pressure them to become ruthless competitors who devote their entire lives to getting ahead, ceaselessly producing and consuming, with no time or appreciation for family or recreation.

Indeed, such a dystopian scenario may not too far from the reality of an increasing number of industries. In such a context, we can argue that negative social pressures are redefining what

the good life entails, and therefore, under the welfarist account, what enhancements are. Here, enhancements appear to be interventions that would improve one's ability to compete within that dubious socio-economic set up. If so, there may be two problems here.

One is that individuals within such a context might be pressured to use interventions aimed at keeping them competitive – if they don't want to use them they will be less able to maintain their current level of well-being, let alone increase it. As for those who want to use these enhancements but can't afford them, this would be unfair to them as well as potentially unjust. However, why is this a problem for the welfarist account? As we've seen, the fact that this account highlights well-being does not mean that it does so at the expense of other values. It just means that these values are not part of the *definition*. If a society is structured in a way so that increasing one's well-being must come at the cost of diminishing another's, then we may be required to restrict access to enhancements. The welfarist account is perfectly compatible with such a conclusion.

The other problem might be that, under such a scenario, social forces have redefined well-being into something that appears to be quite impoverished: the good life here entails becoming a ruthless competitor that spends their life contributing to an ethically suspect economic framework. However, the reason we find this vision of well-being disturbing is because it is difficult to imagine how it can be called a vision of well-being to begin with.⁷ Something has clearly gone amiss. Members of that society are no longer actually working towards increasing their well-being in any plausible sense – rather, they appear to be stuck doing something that actually diminishes it. Or, at least, they are stuck doing something that may significantly diminish their well-being if they stopped. In that regard, this society has simply made an acceptable level of well-being very difficult, and therefore interventions that are purported to be enhancements in that society are not actually enhancing in the welfarist sense.⁸ That is, they are not enhancements at all.

Hence, contrary to the concerns by Camporesi (2014) and Gordon (2014), it's not clear how the welfarist account underestimates the relationship between social forces and one's conception of well-being. Negative social forces can lead to a context where the use of enhancements is unjust, or where the notion of well-being has become highly contaminated and therefore no

⁷ Indeed, under such a scenario a good life might entail becoming a radical non-conformist – to the extent that doing so might lead to a more autonomous, meaningful, or even pleasurable life.

⁸ For instance (and as we will see in chapter five), the fact that smart drugs are sometimes called cognitive enhancers does not mean that their use tends to improve well-being. A society that provides smart drugs as "enhancers" but where those drugs are used to compete in an economy that works to ultimately diminish people's well-being would not make those drugs enhancers in the welfarist sense.

longer related to any plausible conception of a good life. In the former case, the welfarist account is compatible with restricting the use of enhancements, and in the latter, what society deems as purported enhancements may not be welfarist enhancements.

Nevertheless, there is more to be said about the relationship between what might be good for a person, and the social forces that play a role determining that. Examining this issue further will be key to a fuller examination of the account, and so this is the topic for the remainder of this chapter. Before moving on to it, it is worth briefly summarising what we have covered so far.

I have argued that the reasons critics have presented for maintaining the distinction between science-based interventions and conventional ones, and the distinction between therapy and enhancement, are not persuasive. They unhelpfully tangle the discussion with hazy terms that distract from what is more relevant to our deliberations: the well-being of the individual undergoing a modification. Moreover, I argued that while well-being may resist precise definition, it remains useful and open to re-evaluation, with the different theories of well-being converging on a host of goods deemed central to a good life. The welfarist account also does not prematurely settle the ethical issues of enhancement: while it focuses debate on the nature of well-being, it does not do so in an exclusive way, allowing for other values to also determine whether an enhancement is permissible. Moreover, while social forces may negatively influence how we define or pursue well-being, this is not a problem for the welfarist account: such an influence will mean that either using enhancements should be restricted because they lead to injustice, or enhancements are no longer possible as leading a good life is not compatible with the workings of that society. Altogether, the key implications of these conclusions are as follows: enhancement is a much broader concept than the literature takes it to be, and it is the nature of well-being and how it is affected by bodily and mental interventions that should occupy the bulk of our deliberations in this domain.

4. Enhancement, disability, and the riddle of the relevant circumstances

The remainder of this chapter focuses on the relationship between what might be good (or bad) for a person, and the social forces that play a role determining that. Unlike the previous section, where I rounded up objections to the account and responded to them, the following critique

(and the resolution I formulate for it) reflects my own contribution to refining the account and to the debate more broadly.⁹

The key issue here boils down to some neglected implications from the use of “relevant set of circumstances” in the definition of enhancement. As I previously stated, this clause entails that there can be no context-independent answer to the question of whether an intervention is an enhancement. However, very little has been previously said about what this implies. As we will see, it can entail that the account would label certain objectionable changes to a person as enhancing – changes that seem to reflect negative social forces such as racism or homophobia.

To appreciate how that is, we need to first draw on how the welfarist account also defines disability, which will also help sharpen what enhancement is under this account. To that end, suppose you have access to a safe drug that allows you to draft manuscripts with astonishing speed, clarity and insight, while I, after a severe blow to the head, permanently find reading impossible as words now seem to keep jumping across the page. Are these changes to our respective capacities instances of enhancement and disability?

While a functionalist account would judge these changes as enhancing and disabling respectively, the welfarist account gives an unusual answer: it depends. If your new ability to mass produce great manuscripts is likely to make your life go better, while my inability to read is likely to make my life go worse, then yes – you have been enhanced and I disabled. On the other hand, had we both been living in illiterate societies or eras with no use for reading or writing, then it’s not clear there is anything enhancing or disabling about our new capacities. This is because, for the welfarist account, whether a loss or gain in capacities is enhancing or disabling depends on how such changes to our bodies and minds tend to impact our well-being in our given circumstances.

Hence, while enhancement entails a change to a person’s body or mind, an enhanced *state* is:

Any state of a person’s biology or psychology which increases the chances of leading a good life in the relevant set of circumstances.(Savulescu et al., 2011)

Hand in hand with this, the account defines a disabled state as:

⁹ A modified version of this section of the chapter was published in the Journal of Medical Ethics: Zohny, H. (2016). Enhancement, disability and the riddle of the relevant circumstances. *Journal of Medical Ethics*, 42(9). 605-610.

Any state of a person's biology or psychology which decreases the chances of leading a good life in the relevant set of circumstances.(Savulescu et al., 2011)

To return to our topic for this section, what constitutes these relevant circumstances and what makes a given circumstance relevant or irrelevant when determining whether a state is enhancing or disabling? The obvious answer would seem to be that, if our focus is on the well-being of a person, then we ought to include *whatever* circumstances impact a person's well-being. These may be physical, socio-cultural, personal, or anything else conceivably relevant to a person's well-being.

But consider one side effect of such inclusiveness. We may call this the "mislabelling objection"¹⁰ – the key issue that will occupy our remaining concerns. Imagine a person with dark skin in a set of circumstances where there is social prejudice against people with dark skin. If that prejudice reduces their chances of leading a good life, the welfarist account would have us conclude that dark skin (in such circumstances) is a disability. This seems strange; possibly offensively so. Is it relevant or coherent to call that person disabled? For instance, this seems to imply that all black slaves in the Americas were disabled. Moreover, it implies that if a dark-skinned person in a racist society underwent some intervention that gave them paler skin, then, presuming this improves their wellbeing by allowing them to sidestep prejudice, they are now enhanced.

Such a dubious use of labels seems problematic for the welfarist account. Or, at least, in introducing and elaborating on their welfarist definition of disability, Savulescu and Kahane (Kahane & Savulescu, 2009; Savulescu & Kahane, 2011) imply that it is. In fact, as we will see, they hold this as sufficiently troublesome so as to justify excluding social prejudice from the *relevant* circumstances that impact an individual's well-being when defining disability. That is, while the welfarist account holds any given bodily or mental state as potentially enhancing or disabling, if that state or any change to it improves or diminishes a person's well-being solely due to it being subject to social prejudice, then this is considered a separate matter. The impact of social prejudice on well-being is simply not relevant to questions of enhancement and disability. We ought to deal with social prejudice by combating it, not by allowing it to determine whether someone has bodily or mental states that are enhancing or disabling. Or so Savulescu and Kahane (Kahane & Savulescu, 2009) have argued.

While this may appear to sidestep the mislabelling objection, I will argue that it also defeats the purpose of the welfarist account. As I have argued, part of the attractiveness of this account is

¹⁰ I thank Mike King for this way of phrasing it.

in its ability to avoid a number of the conceptual difficulties facing other definitions of these concepts. It does so, I will elaborate, by offering a non-moralized, prudential account of them. However, by excluding social prejudice from the relevant circumstances, I argue the account ultimately loses this distinguishing feature. To state the argument to be presented simply: in trying to avoid the mislabelling objection, Savulescu and Kahane end up retreating to the very position that the welfarist account was designed to avoid. To see how, let's explore what that position is and how the welfarist account is meant to evade it.

4.1 The conceptual link between enhancement and disability

As we have seen, the primary concern of the welfarist account is well-being – or, more precisely, how a specific individual's biological and/or psychological states affect their chances of leading a good life, which I take to refer to their well-being. It is the likely impact of those states on a person's well-being that determines whether they are advantageous or disadvantageous, enhancing or disabling. In other words, the primary concern here is what is of prudential value to a person.

As Savulescu and Kahane (2011) make clear:

If something leads to a reduction in someone's wellbeing, then that thing is bad for that person. This means that if something is a disability in our [welfarist] sense, then it is also a condition that makes life worse.

Likewise, we might add, if something increases someone's well-being, then that thing is good *for* that person. Hence, enhancements here are *any* changes to one's biology or psychology that tend towards that end: increasing their chances of leading a good life.¹¹ What this implies is that enhancements are simply alleviations of disabilities – that is, alleviations of bodily and mental states that are obstacles to greater well-being.

This point tells us what the account holds a person enhanced or disabled *relative to*: a person does not become enhanced or disabled relative to their previous capacities, and certainly not relative to the typical functioning of their species. Instead, they are enhanced or disabled relative to how well their life could possibly go given certain bodily and mental states.

¹¹ Why the emphasis on "any"? Again, this points to the focus on prudential value here. It is not relevant what the biological or psychological state is or what the nature of the change to that state might be (e.g. drug-based or education-based). What matters here is how that state (or how a change to that state) impacts well-being.

This has a further counterintuitive implication: we are all disabled all of the time, and to varying degrees. This is because we all arguably possess mental and bodily states that get in the way of us leading better lives. Hence, even a highly enhanced “posthuman” may nevertheless be disabled on this account.

Unusual as this may seem, the key here is to remember that this is a revisionist account of enhancement and disability as concept: the account stems from a premise that our more common conceptions of these terms are inadequate and in need of upgrading if they are to capture what is of relevance to the nature and ethics of enhancement and disability.

To see how inadequate these more common conceptions are, note how they are typically expressed. As stated, enhancement is most often characterized in functionalist, transmedical terms as an intervention that augments our capacities in some way (Bostrom & Roache, 2008; Engelbart, 1962), or more specifically, as an intervention that does so in a manner that goes beyond merely restoring health or preventing illness (Daniels, 2000; Juengst, 1998; Pellegrino, 2004; Resnik, 2000). In contrast, disability is often broadly understood in medical terms as some unfortunate deviation from species-typical functioning (Areheart, 2008; Eichhorn, 2008; Silvers, Wasserman, & Mahowald, 1998).¹²

While the conceptual problems facing these accounts are numerous (Earp et al., 2014; Kahane & Savulescu, 2009; Savulescu & Kahane, 2011; Savulescu et al., 2011), what unites them is their failure to say anything about how disabled and enhanced states relate to well-being. As I previously pointed out, on these functionalist accounts, a person who augments their auditory capacities so they can hear everything within a mile has been enhanced, *even* if such a capacity leaves them chronically distracted and ultimately deeply neurotic (and therefore, we can assume, unable to lead a good life). Similarly, on these accounts, a person who falls below species-typical functioning by becoming, say, deaf, is considered disabled, even if they are leading a life of much higher well-being than they otherwise would have due to, say, their enthusiastic immersion in a rich Deaf culture, their enjoyment of a heightened sense of touch, along with the fact that they simply have no desire to hear.

More importantly, these accounts raise the question of what the moral relevance is of standards like species-typical functioning or “normal” functioning. By defining these terms relative to one’s prior state of health or one’s species-typical functioning, the implication is that these standards

¹² There is also the social model of disability, which sees disability as resulting not from any impairment or dysfunction, but from an unjust failure of society to accommodate differently-abled or impaired individuals. I will examine how this model relates to the welfarist account in a later subsection.

serve as morally-relevant reference points in and of themselves. And yet, without explicating how these reference points or boundaries are relevant to well-being, or at least some other uncontroversial value, their usage in characterizing enhancement and disability reflects an arbitrarily moralized stance on the matter. On the other hand, what is the moral relevance of people leading better or worse lives? For some ethical theories, that is all that morally matters, while for other ethical theories, that is at least a cornerstone of what is relevant to ethics.

As such, by tying these concepts to well-being, the welfarist account allows us to avoid the moralized starting points and terminological controversies at the core of other accounts. It provides an inherently normative approach, giving us *prima facie* reasons to address bodily and mental states that are making an individual's life worse, as well as helping us understand why enhanced states can be desirable to begin with (Savulescu et al., 2011). This sets the groundwork for a germane and coherent framework from which to then think about what is at stake in regards to enhancement and disability without having to rely on controversial references to "normality" or other morally arbitrary standards of functioning. In that way, the welfarist account sidesteps these pitfalls and keeps the focus on what actually matters in a given case: a person's well-being.

Nevertheless, what remains troubling about this account is its emphasis on the "relevant circumstances", and here we home back in on the central issue. To start us off, why does it place such emphasis on a person's circumstances in its definitions of enhanced and disabled states?

The reason is straightforward: what makes leading a good life easier in one circumstance may make it harder in another. One example Savulescu and Kahane (2011) give is our disposition to rest and store excess calories as fat: such a disposition is clearly advantageous in a set of circumstances where resources are limited, but it can lead to ultimately lethal consequences where there is an over-abundance of resources. Even something as extreme as a brainstem stroke that leaves an individual unable to speak or move – most would agree that such a condition severely limits one's chances of leading a good life, yet surely that is true *only* in relation to our broader circumstances.

What this shows is that, as I've argued, there is no context-independent answer to the question of whether a state increases or decreases the chances of leading a good life. The particular circumstances of the individual (which may include everything from the type of universe they inhabit, to their surrounding culture, to their personal life goals and desires) clearly play a determining role in answering that question.

And yet, this brings us back to the central conundrum raised above: a person with dark skin whose set of (in this case, social) circumstances include prejudice against dark skin. The welfarist account may label that person as disabled, which seems intuitively wrong. Let's explore the nature of this mislabelling objection in more detail, including what it might mean for enhanced states.

4.2 The mislabelling objection

Again, consider a person with dark skin who lives in a racist society that diminishes their wellbeing by discriminating against dark skinned individuals. Moreover, imagine this person hears of a new intervention that promises to make them significantly paler. This need not be a speculative technology: skin whitening creams exist, and advertisements for them have played a contentious role in India (BBC, 2003). Our imagined person does not want to spend their lives battling unjust discrimination, so they agree to undergo or use this intervention, and as a result are not discriminated against as often, increasing their prospects and income, and ultimately their level of wellbeing.¹³

According to the welfarist account, this person has been enhanced. Indeed, since enhancements on this account are alleviations of disabilities, this person was disabled: they possessed a biological state or property (dark skin) that reduced their chances of leading a good life in their circumstances. Similarly, the account suggests that the homosexual who becomes heterosexual in a homophobic society, or even the woman who reassigns her gender in a misogynistic one, have become enhanced, if they are prudentially better off because of the change. Furthermore, those who do not undergo these modifications in such contexts may be said to live with a disability.

Or consider societies where girls who have not undergone some form of female genital mutilation (FGM) are discriminated against, so much so that (we will presume) the loss of wellbeing from such a practice is outweighed by the loss of life prospects and basic, decent treatment by other members of her society. The welfarist account may have us label women

¹³ It is possible that some accounts of well-being would not consider this a genuine increase in wellbeing. Perhaps an objective list theory of wellbeing might classify conceding to prejudice to be incompatible with leading a good life, at least to some degree. However, objective lists tend to take into account pleasure and autonomy. It is uncontroversial to assume that hedonistic or desire satisfaction accounts of well-being may consider this an instance of genuine increased wellbeing.

who have undergone FGM as “enhanced”, with new interventions that facilitate that practice as “enhancement technologies”. Indeed, according to the welfarist account, having a clitoris in such circumstances may count as a disability.

In these scenarios, the use of these labels seems troubling. On the one hand, the mislabelling objections acts to pump an intuition that something has gone amiss here – the account has simply lost the plot in an important respect. And there may be good consequentialist reasons for thinking that. Superficially at least, there appears to be something politically incorrect here. Calling a pale skinned person or a straight person “enhanced” compared to being dark skinned or gay – some may wish to argue that such labels would be unacceptable. And one reason for that might be that the use of these terms may act to validate or play into the already existing injustice within those societies. For instance, by labelling a girl who has had her clitoris removed “enhanced”, there may be a concern that such a positive label could reinforce the social forces behind the practice. Similarly, the “disabled” who fail or refuse to undergo “enhancement” might risk further discrimination thanks to the stigmatizing connotations that may be associated with the mere label of “disabled”. This may include subtler forms of discrimination such as aversive disablism: an unconscious prejudice against individuals considered disabled due to a belief that they are inferior to others (Deal, 2007).

Moreover, there might be a concern that controversial enhancements on this account could be more easily assimilated as therapies within medicine. For instance, if being gay can be coherently labelled a disability – and it *can* be coherently labelled so on the welfarist account, given the right circumstances – then this may act in the favour of medicalizing being gay and offering biomedical interventions to somehow curb it (on the efficacy of “conversion therapies” see Spitzer, 2003). Indeed, it may help medicalize whatever biological or psychological states that victims of prejudice possess, further propagating the notion that they embody some lamentable deficiency (Titchkosky, 2001).

Nevertheless, the proponent of the welfarist account has a straightforward response to these concerns. Firstly, we should remember that this account is not trying to define what types of biological or psychological states are inherently better. Rather, it is trying to capture what may be better or worse *for* a person in *their* circumstances – that is, what may be advantageous or disadvantageous for them. And, it is a matter of fact that being, say, gay or dark skinned for some people in some circumstances is worse *for* them than being straight or white. It is purely this prudential dimension of their states that this account strives to capture. And while it may be true that there are negative connotations that follow from the label “disabled”, this is only

due to an understanding of disability as some morally relevant deviation from “normality”, or as some condition warranting medical attention. Yet it is precisely such an understanding of disability that the welfarist account seeks to move us away from.

Furthermore, if the interventions discussed here (e.g. somehow changing a person’s sexual orientation to a conforming standard) were deemed to promote injustice, the welfarist account is compatible with regulating those interventions in whatever way we deem just, including banning them. The fact that an intervention may be of prudential value to someone does not in itself entail that it ought to be permissible. This is because, as I have argued throughout this chapter, the welfarist account is not a stance on the permissibility of enhancement, it is a rather a definition of the concept. While it highlights well-being as a starting point for thinking about enhanced and disabled states, how and when we decide to promote an individual’s well-being would also be determined by other values such as justice, as well as whatever other values individuals or society think are relevant to ethical decision making (Savulescu et al., 2011). As I argued above: nothing about the welfarist account entails that the promotion of individual well-being trumps all other values, let alone that such a promotion should be realised through biomedical means, as opposed to environmental changes. I explore in more detail the relationship between prudential and moral value, and its implication for the welfarist account, in chapter four.

4.3 Excluding social prejudice (and why this fails)

Despite this possible response by the proponent of the welfarist account, Kahane and Savulescu take a different track (Kahane & Savulescu, 2009; Savulescu & Kahane, 2011). As I mentioned previously, they attempt to sidestep this issue of alleged mislabels by excluding social prejudice from their definition of disability – all at the expense, I argue, of robbing the account of its key advantage over other definitions. How do they do this?

While they define disability broadly as any state of a person’s biology or psychology which decreases the chances of leading a good life in the relevant set of circumstances (Savulescu et al., 2011), their more detailed analysis of the concept includes the addition of this clause: “excluding the effect that this state has on well-being that is due to prejudice against that person by members of that person’s society” (Kahane & Savulescu, 2009; Savulescu & Kahane, 2011).

Let us call the addition of this exclusionary clause to the welfarist definition of disability “W2” – to be contrasted with W1, the standard account.

What W2 entails is that if a biological property or state is reducing the well-being of a person solely because that state is an object of prejudice in their particular social circumstances, then that state does not constitute a disability. That is, W2 redefines disability as the effect a bodily or mental condition has on well-being that results *after* we subtract the effect it has due to the effects of social prejudice (Kahane & Savulescu, 2009). This means that, in the case of the dark-skinned person, or the homosexual, or the non-mutilated woman in societies that are prejudiced against them – these are victims of prejudice, not disability.

Indeed, Savulescu and Kahane adopt W2 precisely as “it relieves us from having to claim that having dark skin colour in a racist society is a disability” (Kahane & Savulescu, 2009).

To more fully grasp what W2 entails, consider a more complex situation involving an individual bound to a wheelchair in a society that does little to accommodate them due to prejudice against individuals in wheelchairs. Here W2 would hold that, if this individual is disabled, it is not because of their loss of well-being due to social prejudice against them, but because their condition leaves them unable to, say, fully enjoy the benefits and/or joys of bipedal mobility. These are benefits that, even if their society had fully accommodated them, they still would not have access to (at least given current levels of technology). On the other hand, the added diminishment to their well-being from social prejudice against them – that does *not* form part of their disability. It is a separate issue that calls for a different response: to combat that social prejudice. Hence, an individual may be disabled *and* be a victim of prejudice against their disability, but that prejudice does not constitute part of their disability.

But what might the implications of W2 be on the welfarist definition of enhancement or enhanced states? While Kahane and Savulescu do not discuss this, given that enhancements act as alleviations of disabilities, excluding social prejudice from the definition of enhancement would entail that a person with dark skin in a racist society who undergoes an intervention to have paler skin (and thereby, we will presume, improves their well-being) does not become enhanced; they were never disabled to begin with. Rather, they were affected by prejudice, not disability. The modifying intervention has merely helped them sidestep that prejudice, but it has not enhanced them in the way relevant to what the addition of W2 makes the account now tries to capture: changes to our biology or psychology that improve our well-being for reasons unrelated to social prejudice.

At this point, however, it is starting to become clear the welfarist account is no longer purely concerned with the prudential value of a person's bodily and mental states. Like the other accounts it critiques, it too has become moralized, making a value judgement on the sorts of circumstances that are relevant to whether states in our bodies and minds are disabling or enhancing. On the one hand, this value judgement may allow us to characterize disabilities in a more specific way. For instance, imagine being told of an old friend who has immigrated to another country only to become disabled soon after. You then inquire about the nature of their disability, and are told that they had fallen victim to racism. Here, the term "disabled", even if understood as a reference to well-being, seems unhelpfully vague and distracts from what we might want to say is the real issue: racism. But by excluding social prejudice as W2 does, we can use the term to distinguish cases like these from ones where the nature of someone's disability is rooted in blameless forces, like physical laws and/or biological facts, as well as non-discriminatory social forces (Kahane & Savulescu, 2009). This, it seems plausible, is a useful and morally relevant distinction that our concepts should be able to make.

And yet, if it is only useful or morally relevant distinctions that we are after, surely there are many more to be made. After all, this account is concerned with *any* biological or psychological state, hence whether one becomes, say, mildly cognitively impaired (MCI) due to head trauma or due to old age is irrelevant. According to the welfarist account, including W2, both are, or can be, disabling. But most of us would think there is a relevant distinction here. For instance, MCI due to a head trauma at a young age tends to be perceived as tragic and warranting treatment in a way that "natural" MCI in old age is not and does not. So why not add a further clause to exclude well-being-diminishing states that we might deem 'natural', such as MCI with old age? That would make the account more in line with our intuitions about what is relevant to disabilities. In fact, in regards to enhancements, why not add a further clause that excludes changes to the body or mind that may exacerbate injustice? Surely there is a relevant moral distinction between an enhancement that comes at little cost to others in society and ones that cause great societal harm (by, for instance, increasing discrimination against those who refuse or can't afford to undergo the intervention).

The answer to these questions lies in what the welfarist account is concerned with highlighting: a person's well-being, not the morality of the social circumstances that increase or diminish it. Again, it is designed to capture what is advantageous or disadvantageous for a person in their circumstances. Whether certain states *ought* to be advantageous or disadvantageous may be highly relevant questions to ask, but these would be ethical questions about prudential value: when it ought to be pursued, and how, and in what ways it ought to be distributed. Yet the

welfarist account is explicitly formulated to help us non-judgementally answer the question of what states are good or bad for a specific person in a specific circumstance. Once we open the door to moralizing this account, as this exclusionary clause in W2 does, this raises the question of why we should stop there and not add other boundaries to it. Indeed, it raises the question of why we should even consider this account over the other, more common, definitions that appeal to species-typical functioning or other morally arbitrary reference points.

Also, the issue of vagueness raised in the case of the old friend who has immigrated is not as problematic as it may seem. This is because even the everyday language use of the term “disability” tells us little on its own about the nature of any given disability. Is it congenital? Is it due to an accident? How debilitating is it? The answers to such questions may have substantial consequences for how we respond to a given disability – for instance, whether we decide to treat it, and with what degree of priority, or whether we ought to change how our society is organized to accommodate it. Without these details, the label “disability” tells us very little. With that in mind, the welfarist definition would not make the concept any more vague in regards to a given individual. If we accept the welfarist account, we merely have to add other equally relevant questions to the ones above: Is it due to old age? Is it due to social prejudice? Indeed, is it due to non-discriminatory social forces? It’s not clear that a disability arising due to social prejudice would be unique in its impact on how we might perceive or respond to a person suffering from it.

Interestingly, despite adding this exclusionary clause, Savulescu and Kahane do not deny that social prejudice is relevant to disability. They present their welfarist account as an approach that occupies a middle ground between the two most common ways of understanding disability: medical and social models of disability. The medical model typically characterizes disability as a stable property of a person that deviates (or falls below) the normal functioning of the human species. It reflects some deficiency or lack in motor, sensory or cognitive capacities that most people possess, implying that such a deficiency is also a misfortune that ought to be corrected if possible (Areheart, 2008; Eichhorn, 2008; Silvers et al., 1998). In contrast, the social model does not see disability as an impairment in itself, but as the disadvantage or restriction of activity caused by a contemporary social organization that takes little account of people who have physical impairments (UPIAS, 1976). Hence, for the social model, it is the unjust way through which society is organized that is disabling for some (Amundson, 2005; Koch, 2001).

The welfarist account of disability seeks to occupy a middle ground between these two approaches in this way: like the medical model, it suggests that disabilities refer to conditions or

states that an individual possesses. Like the social model, however, it denies that those conditions or states are, in and of themselves, morally significant or unfortunate. On the other hand, unlike the social model, the welfarist account disagrees that the only way disabilities reduce well-being is through social prejudice. But this is not to say that social prejudice is unrelated to disability. As they make clear:

Advocates of the social model are no doubt right that some of the adverse effects of deviation from the species norm are due to such prejudice. We disagree, however, that all of the ways in which common disabilities can reduce wellbeing are entirely due to prejudice. (Savulescu & Kahane, 2011)

This suggests that the disagreement here is not that social prejudice is a separate matter to disability, but only that there is more to disability than mere social prejudice. Given such a concession, it seems strange that Kahane and Savulescu then go on to entirely exclude social prejudice from their definition by adding W2 to it. The definition now appears to shun the very essence of the social model. W2 can, in principle, count *any* abiding condition as a disability, but not the one condition that, for the social model, is *the* condition: social prejudice. In that way, the welfarist account no longer seems to occupy a middle ground between the medical model and the social one. It now fails to capture the fact that disability can arise from how, both, discriminatory and non-discriminatory circumstances interact with our biological and psychological states to impact our well-being positively or negatively. More importantly, it now retreats to the very position it is designed to avoid: a moralized account with (what I have argued is) an arbitrary boundary built into its definitions.

4.4 An alternative to excluding social prejudice

Is there a better alternative to avoiding the mislabelling objection, aside from excluding social prejudice? I think there is. It lies in the fact that this account seeks to capture the nature of advantageous and disadvantageous biological and psychological states. Indeed Kahane and Savulescu want to use these terms interchangeably with being enhanced or disabled (Savulescu et al., 2011).

Given this intention, note how if we were to rely solely on the term “disadvantageous states” instead of disability, the mislabelling objection seems to vanish. This is because, while it may seem perverse to call a dark-skinned person in a racist society “disabled”, it seems far less controversial to claim that, in a circumstance where social prejudice entails that having dark skin

is disadvantageous, it *is* disadvantageous to possess dark skin. Having dark skin within such circumstances is a biological state that is disadvantageous precisely because it reduces the chances of leading a good life. Similarly, being gay in a homophobic society can be a disadvantageous state to the extent that it reduces the chances of leading a good life within such circumstances. It is coherent to say that being gay can be a disadvantageous state, but that it shouldn't be. Such a description does not have the potentially discriminatory or medicalizing connotations that arise when we use the label "disabled" for individuals in these situations. On the contrary, when society leaves people at a disadvantage because of prejudice, the connotation is that society should do better for them, not that they suffer from an unfortunate condition that ought to be "fixed".

Similarly, if one were to undergo a biomedical intervention that allowed them to sidestep social prejudice, we can say that they now possess more advantageous biological or psychological states. They are advantageous *for them* precisely because they enable them to sidestep that prejudice. And so, while it may seem wrong to call a dark-skinned person who becomes paler in a racist society "enhanced", it is uncontroversial to argue that their paler skin is an advantageous biological state to possess in their particular circumstances, even if it is unfair that it is advantageous.

Based on this, if the welfarist account eschews terms like enhanced and disabled and adopts advantageous and disadvantageous states, it can sidestep the mislabelling objection, and there would be no pressure to add a clause excluding social prejudice from its definition of disability. The benefit of this is that it would enable the account to achieve what it sets out to achieve: to acknowledge that prejudice can play a role in making a certain biological or psychological state disadvantageous, but that disadvantageous states can certainly also arise for reasons unrelated to prejudicial social circumstances. In that sense, it can indeed occupy a middle ground between the social and medical models of disability, while also avoiding the retreat to being a moralized account.

But how satisfactory would such a substitution in terms be? While it may offer a simple solution, in practice it is arguably unlikely to catch on in the voluminous disability and enhancement literature, let alone in everyday speech. A compromise may be to increasingly interchange our use of enhanced and disabled with advantageous and disadvantageous states, until these acquire a synonymous relationship. That, at least, is what I will be doing for the rest of this thesis.

Here, we would not necessarily be biting the bullet on the mislabelling objection when we do periodically use the terms enhanced and disabled. Recall that the welfarist account seeks to

undermine the notion that disability is a (necessarily) morally significant deviation from normality, or that disability entails a medical problem. It is therefore working to neutralize the negative connotations and societal attitudes that often accompany the label of being disabled. The more it is adopted, the less likely concerns related to the mislabelling objection would materialize.

Of course, the fact that the account seeks to undermine these negative connotations of disability does not mean it will succeed. It may be that even a well-being-grounded expansion of the range of states that we consider disabling will nevertheless lead to prejudice against people with those states in the form of aversive disablism. That is, even with an increasingly accepted welfarist understanding of disability, we may still find that some will subject others who are understood as disabled to subtle forms of prejudice due to an entrenched notion that the disabled are inferior to others. Whether this will likely happen if we adopt the welfarist account and use the term “disabled” may be posed as an empirical question that invites investigation. Depending on the answer, we may wish to more confidently embrace the terms enhanced and disabled in the welfarist sense, or we may have to move away from them and settle for advantageous and disadvantageous states instead.

Either way, we can end by reiterating that the welfarist definitions of enhanced and disabled states are revisionist: they stem from a premise that our common conceptions of these terms are inadequate and in need of upgrading. If we do use the terms enhanced and disabled, we may have to accept that, in some situations, we will (perhaps somewhat awkwardly) label individuals as disabled or enhanced, despite the fact that this feels like an alien or injustice-promoting use of these words.

In such cases, we would need to remember that we are using these labels merely to describe states that are good or bad *for* a person in a certain circumstance, not states that are good or bad in and of themselves. In other words, we are merely describing states in terms of their tendency to promote prudential value. But whether a given state *ought* to make a life go better or worse, and whether it ought to be promoted or avoided and in what ways – these are questions to be settled in conjunction with other values, such as justice. The welfarist account offers a solid framework to help us think about these further issues without being hampered by the terminological controversies and conceptual confusions of other accounts – so long, that is, as we resist the temptation of adding the moralizing baggage that comes with excluding social prejudice.

5. Conclusion and further thoughts

The welfarist account has been criticised for obscuring the distinction between science-based interventions and so-called conventional ones, as well as the distinction between therapy and enhancement. It has also been objected to on the basis that it relies on the contested concept of well-being; that it is unhelpful in classifying interventions; that it appears to settle ethical questions about enhancement through some philosophical sleight-of-hand; and that it underestimates the role social forces play in how we define well-being, as well as the collective consequences of enhancement.

I have argued that these criticisms are either unfounded or addressable. The distinction between science-based interventions and conventional ones, and the distinction between therapy and enhancement, are not useful, only serving to distract from what is actually of relevance: the well-being of the individual undergoing a modification. Moreover, while well-being may resist precise definition, it remains useful and open to re-evaluation, with the different theories of well-being converging on a host of goods deemed central to a good life. The welfarist account also does not prematurely settle the ethical issues of enhancement: while it focuses debate on the nature of well-being, it does not do so in an exclusive way, allowing for other values to also determine whether an enhancement should be permissible. Moreover, while social forces may negatively influence how we define or pursue well-being, this is not a problem for the welfarist account: such an influence will mean that either using enhancements should be restricted because they lead to injustice, or enhancements are no longer possible as society has made leading a good life impossible. Finally, I considered the account's emphasis on the relevant set of circumstances and the mislabeling objection this seems to raise. I argued that becoming clearer about the prudential focus of this account highlights how this is ultimately an account of context-dependent advantageous and disadvantageous bodily and mental states. Understood as such, the mislabeling objections loses its thrust.

At this point, it is important to link back to my primary thesis goal: providing a coherent account that, if shared, can help clarify our thinking about enhancement and allow practical ethicists and policy makers to more pragmatically deliberate about and respond to putative enhancement technologies. In the introduction to this thesis, I argued the welfarist account can arm bio-conservatives with a point of focus to their arguments that we can all agree matter: the impact of interventions on well-being. I noted that most objections to enhancements in the functionalist, transmedical sense are in fact concerns about their impact on our ability to lead

good lives (they may rob us of our ability to lead authentic, autonomous, or meaningful lives.) By voicing these concerns in welfarist terms, we can move the discussion somewhere more productive: not whether an intervention is permissible or not, but what a good life is, and how an intervention is likely to contribute to or detract from it (I return with a concrete example of doing this at the end of chapter five). And as for objections that are not rooted in a concern for the well-being of people who use a transmedical enhancer, but in the injustice or unfairness for others that this may lead to, I hope it is clear how the welfarist account can also facilitate clarifying these objections: the fact that something is an enhancer does not mean it is permissible, all things considered. What the welfarist account does is help us make this distinction. We can agree that an intervention will likely be advantageous to a person, but raise other concerns regarding the impact on others. The welfarist account is compatible with this, and for those deploying it, clarifying their objection against (or promotion for) a certain intervention in this way can lucidly move discussion forward. As I stated, we may well find that where we disagree is on what leading a good life is and what justice is, rather than whether a certain class of intervention is permissible or not.

Ultimately, while the welfarist account entails much philosophical work (namely, what is a good life and in what ways should we best pursue one), it also calls for much empirical work. The number of interventions being called enhancers is growing, especially amongst the off-label use of certain drugs and brain stimulation techniques. The welfarist account suggests that we now have to ask if these putative enhancers improve the well-being of their users. That is, it impels us to ask if they actually are enhancers. I undertake an example of such a task in chapter five. However, it raises other questions too: so far I have presented it as a general account of enhancement, but can it account for moral enhancement? Its focus on prudential value suggests a conflict, potentially limiting the scope of the account. I explore this topic in chapter four. First, let us further investigate this account's referral to a person's psychology and what this might entail for enhancement. As I noted in the introduction to the thesis, while what a person's biology entails is clear enough, the extended theory of the mind suggests our psychologies may literally be constituted in part by our environments, potentially further expanding the scope of certain forms of enhancement.

Chapter 3

Enhancing the Extended Mind

This chapter focuses on the scope of what is entailed by changes to a person's psychology. The welfarist definition of an enhancement is any change in the biology or *psychology* of a person which increases their chances of leading a good life in the relevant set of circumstances. In the context of the discussion on enhancement, making a change to one's psychology ordinarily brings to mind interventions like psycho-active drugs, or perhaps some form of brain stimulation. To the extent that those impact the mental characteristics of a particular person, we can say they have changed their psychology. And if that change (or its repeated occurrence) tends to increase the chances of that person leading a good life, the welfarist account deems it an enhancement.

However, note the internal nature of these changes: they are changes that occur through some impact that takes place *inside* our bodies or, more specifically in this case, our brains. However, in this chapter, I argue that, when it comes to our psychology, it is not clear that it is constituted purely by structures and activities internal to our bodies or brains. In fact, our psychology – and specifically our cognitive processes – in a sense spills out in the surrounding environment, and changes to that environment can literally constitute enhancing (or disabling) changes to our psychology.

This, at least, is the conclusion that follows from what some have called the extended mind thesis (Clark & Chalmers, 1998). In the first section of this chapter, I sketch out this theory and its implications, and argue it may significantly expand the scope of what counts as changes to a person's psychology, and therefore what might be considered enhancements. In the second section, I consider a crucial objection to the extended mind thesis and argue that it fails. Finally, I explore Neil Levy's (2011) Ethical Parity Principle, which is an argument for why alterations to external props that are part of an extended mind are, all else being equal, ethically on par with alteration to the brain. I consider some challenges to this principle, and argue, following Danaher (2015), that we may have good reasons to favour enhancing cognition directly through alterations to the nervous system rather than indirectly through alterations to external cues and tools.

The importance of this argument for my thesis are two-fold. Firstly, it explores yet another dimension of the welfarist account that has not received attention in the literature – its original formulators (Savulescu et al., 2011) do not specify what is entailed by a person’s psychology, nor have any detractors of the account raised this question. In that sense, this chapter sees how far we can take the implications of the account. Secondly, if this argument is correct, it should factor in our deliberations about how to regulate enhancements. Policies that seek to regulate “internal” enhancements differently simply because they are internal will have to do more to justify such an approach. As we will see, such wariness of internal changes may be harder to justify than seems.

1. Where do mental states happen?

As I have argued, the welfarist account does away with certain contestable dualities inherent to other definitions of enhancement. Namely, the emphasis on *any* change in the biology or psychology of a person entails there is no need to distinguish between therapy and enhancement, nor so-called science and technology-based interventions and more conventional ones.

However, there is a third duality the welfarist account appears to retain. Enhancements can be described as internal or external (Buchanan, 2011). An internal enhancement entails a certain invasiveness into the body (whether a drug, electric current or surgery – or even the neural changes associated with education), while external enhancements relate to the utilization of tools (like cars, calculators, and so on). Having a car can certainly augment the speed at which people can move, but it does so through the use of something external to the person.

The welfarist account, like other definitions of enhancement, focuses on internal enhancements: “Any change *in* the biology or psychology of a person”, as opposed to “any change to a person’s life.” This emphasis on what is internal is important if we are to maintain a distinction between enhancement technologies and technologies in general or the general socio-economic and cultural forces that impact a persons’ well-being. But, as we will see, in some cases, even this duality loses its relevance. This is because a person’s psychology, I will argue, is not limited to internal processes within the brain, or the body of a person.

This is the idea behind the extended mind thesis (EMT). To examine it, let us be clear about what cognition refers to. In the context of cognitive psychology, cognition is the mental process of

perceiving, learning, remembering, and thinking about information (Sternberg, 2011). It thus includes mental abilities such as attention, memory, judgement, reasoning, and so on. These may operate at a conscious or unconscious level. For changes to those processes to be enhancing in the welfarist sense, they must tend to improve our chances of leading a good life. However, the key point here is that, regardless of ongoing disputes about the exact nature of these processes, one uncontroversial consensus is that they are a function of the brain.

However, this raises an interesting question: does that mean they cannot be a function of things other than the brain? Several theories of cognition contend they can. For instance, embodied cognitive science holds that cognition is fundamentally dependent on aspects of an agent's body outside of the brain (Wilson & Foglia, 2011). Mental life as we know it would not be possible without the involvement of the body in acting and sensing. Similarly, the theory of embedded cognition holds that cognitive processes depend very heavily on props and devices in the surrounding environment (Rupert, 2004).

For our purposes, we are interested in a more radical characterization of cognition, one claiming that not only can cognitive states and processes *depend* on the body and objects in the surrounding environment, but they can literally be *constituted* by things outside the body.

This is the central claim of EMT¹. Its earliest defenders, Andy Clark and David Chalmers (1998), illustrate the thesis at work through the case of Inga and Otto. Inga hears of an exhibition at the Museum of Modern Art. Wanting to go see it, she recalls it is on 53rd street, and heads off to that direction. Her belief about the museum's location was stationed somewhere in her memory, waiting to be accessed.

In contrast, consider the case of Otto, who suffers from dementia. Knowing his condition makes him forgetful, he always carries around a notebook, which he regularly updates, and which contains all the information he might need to remember on any given day. If he needs to recall something, he looks it up in his notebook. In that sense, the notebook plays the role that Inga's biological memory played. In his case, hearing about the exhibition and deciding to go, he accesses his notebook for the address, uncovers that it is on 53rd street, and heads that way.

Clark and Chalmers argue there is nothing fundamentally different between Inga and Otto. They both remember the location, except Otto uses something external to his brain for the process of recollection.

¹ Also called environmentalism (Rowlands, 1999), locational externalism (Wilson, 2004) and the Hypothesis of Extended Cognition (Rupert 2004).

At this point, it may be tempting to say that Otto's inability to access this information directly reflects that, unlike Inga, he does not actually hold a belief about the museum's location – he merely holds a belief about where to find that information: his notebook. But consider the role beliefs play. We can define a belief as cognitive content that is held to be true (Schwitzgebel, 2015). Presumably, what makes the content cognitive in nature is that it was acquired at a previous time, processed in some way, and stored. We know that Otto regularly updates his notebook. We also know that, at some point, he processed information about the museum's location and stored it. The fact that the location of the stored cognitive information is outside the brain is not relevant. Saying he only has a belief that his notebook contains the information about the museum's location is like saying Inga only has a belief that her memories contain information about the museum's location. It is true they both possess these beliefs, but it is also true that Otto's notebook entry fundamentally functions like Inga's memory. It is literally a belief, albeit one that resides outside his brain.

None of this suggests that anything in the environment can form part of an extended mind. Certain conditions must be met for this to happen. Hence, if Otto uncovered the museum's address by stumbling on a random, crumpled piece of paper on the street that had "Museum of Modern Art = 53rd Street" written on it, it would not constitute part of his mind. As Clark and Chalmers (1998) highlight, for an external prop to play that role, it must be readily accessible, and its contents must be automatically endorsed and have been endorsed in the past. A random piece of paper is obviously not readily accessible. It is coincidental that Otto found it. Moreover, Otto has no reason to endorse its content. Unlike his notebook, which he knows is reliable and has endorsed in the past, he has no reason to believe the information it contains is correct.

However, if these conditions are met, we have reason to consider an external prop as part of the mind. This idea that such external props can constitute part of the mind is captured by the so-called parity principle. The principle states that, if something plays a role in cognitive activity, such that, were it internal we would have no difficulty in concluding that it was part of the mind, it should be counted as part of the mind whether or not it is internal (Levy, 2007) .

The parity principle works to undermine the notion that there is something special about the brain or the biological matter we associate with cognitive processes. In that sense it is rooted in a functionalist understanding of the mind. We can define functionalism as the view that what constitutes a mental state or process is not its internal constitution (e.g. neurons), but the causal role it plays within the system it is part of (Levin, 2013). The parity principle thus asks whether an extended process functions in a way that, were it taking place in the brain, we would have no

hesitation recognizing it as cognitive (Walter, 2010). Hence, given that Otto's notebook plays a role that, were it occurring intracranially we would have no hesitation identifying it as a cognitive function, the parity principle compels us to consider it as literally part of his mind, or more specifically, a constituting part of his mental processes.

Menary (2006) articulates the parity principle more simply: When met with an external process that, were it located in the skull we'd call it a cognitive one, we ought to call that external process a cognitive one.

Of course, Otto's case is only illustrative of a far more common phenomenon: cognitively healthy individuals using technological props to (sometimes profoundly) augment their cognitive powers. The case in point is the increasingly sophisticated smartphone. Readily accessible, it provides users with information ranging from the exact location of everything known to, say, Google maps, to the entirety of human knowledge as stored on the web. Of course, we would not automatically endorse much of that information due to its unreliability. However, neither do we automatically endorse all the information we recollect from the brain. We are often unsure about the truthfulness of our recollections. But consider the regularity with which we automatically endorse uncontroversial facts stored in the online encyclopaedia Wikipedia – facts we have no hope of storing within our biological memories. Or consider how we endorse information about the location of a hotel in a new city we are visiting through GPS, and how we might continue to automatically endorse that information as we walk around or get lost.

While modern technology may present a special case of augmenting cognitive capacities by extending them, the process involved here is not necessarily limited to modern technology. Consider how two individuals who are close to each other can, in principle, come to share a single extended mind. We can imagine that, instead of a notebook, Otto relies on his close and trusted partner to provide him with the relevant information on a given day. That partner's mind could come to constitute part of Otto's extended mind. More generally, we can see how some cognitively healthy individuals may rely on each other to be reminded of important but forgettable things. In that sense, the capacity (and perhaps desire) to extend the mind is an uncontroversial feature of being human, possibly driving the invention of technologies such as literature, arithmetic and the myriad tools that helps us store, distribute, and calculate relevant information (Levy, 2007).

If this is true, the implications for the enhancement of our psychologies, and specifically our cognitive abilities, are significant. They entail that cognitive enhancement, far from being new or controversial, is an ancient and ongoing human phenomenon that rarely raises any ethical

qualms. External props that come to constitute part of an extended mind literally constitute part of a person's psychology. Changes to those props are hence literally changes to their psychology.

To tie this in with the welfarist account of enhancement, if an external prop constitutes part of a person's cognitive processes (and therefore their psychology, taken as a whole), then changes to that prop that tend to increase their chances of leading a good life (in the relevant set of circumstances) are instances of cognitive enhancement. Hence, upgrading one's smartphone may turn out to be an example of welfarist cognitive enhancement. The question, however, is whether such an upgrade is also *morally* no different, in principle at least, from directly augmenting one's brain, such as by taking a pill or implanting a computer chip. Levy (2011) argues there is no difference, at least not in principle. Before we assess his argument, however, it is worth getting clearer about EMT by considering an objection Robert Rupert (2004) raises against it.

2. An objection to the extended mind thesis

Critics of EMT can resist the full implications of the theory. They may concede the crucial role external props can play in some cognitive processes, but insist those props do not come to *constitute* part of a person's psychology. That is, while one can accept the implications of the embedded theory that cognitive processes depend very heavily on props and devices in the surrounding environment, they may still resist EMT.

That is the objection Rupert (2004) poses. Rupert concedes that "we can properly understand the traditional subject's cognitive processes only by taking into account how the agent exploits the surrounding environment to carry out her cognitive work." That is, he accepts the implications of the embedded theory of cognition. He is sympathetic to the claim that the environment plays a crucial role in cognitive processes, and hence we cannot study a person's psychology in a vacuum if we want to truly understand it – but, he contends, none of this entails cognition is not ultimately an intracranial series of processes.

The issue here is whether EMT offers a better explanation of cognition than the embedded theory. If the embedded theory can allow us to explain cognitive processes by highlighting the importance of the environment, then what is added by going further and claiming that cognition is, in part, literally taking place in the environment? That is, Rupert (2004) argues that if EMT does not offer superior explanations for cognitive scientists compared to the embedded theory,

we should resist taking the (seemingly superfluously) radical step toward EMT “by dint of the methodological principle of conservatism”.

Adams & Aizawa (2010) raise a similar objection. They accuse supporters of EMT of holding an external prop or process to be cognitive by virtue of it merely being connected to a cognitive agent. This, they argue, reflects a confusion about what is “coupled together” and what is “constituted together”. The fact that an object or process is coupled in some fashion to a cognitive agent, does not entail that this object or process *constitutes* part of the agent’s cognitive apparatus or cognitive processing.

However, these objections seem to beg the question against EMT: EMT is arguing for a reinterpretation of what was previously considered a coupling relationship. These objections are invoking intracranial processes as the standard against which other cognitive processes’ status is measured, thereby privileging the very idea that the parity principle was created to undermine. The thrust behind these objections presumes a notion of a discrete, already formed, cognitive agent being coupled with an external prop (Menary, 2006), again relying on the kind “cognitive chauvinism” the parity principle argues we should abandon (C. King, 2016).

Nevertheless, it still seems relevant to ask, does EMT offer any superior explanations for cognitive scientists compared to the embedded theory? In regards to ethical implications, it may not matter which of these theories is correct: so long as we recognize the importance of external props to our cognitive abilities (something that both theories recognize), we will have the same attitude towards changes to the mind and changes to the environment (Levy 2011).

However, if this is all we can conclude, then we have no strong reason to think that a person’s psychology is actually changed when we alter the environment – the fact that it depends on it does not mean that it is *part* of their psychology. If so, this weaker conclusion would have no significant implications for the welfarist account – such props would be no different than, say, housing, which may be crucial to an individual’s well-being, but is not literally part of their body.

The embedded theory, however, is vulnerable to what is arguably an absurd conclusion due to its insistence that cognition is ultimately an intracranial process. To see this, we can build on Andy Clark’s (2009) thought experiment about Diva:

Imagine a case in which a person (call her Diva) suffers minor brain damage and loses the ability to perform a simple task of arithmetic division using only her neural resources. An external silicon circuit is added that restores the previous functionality. Diva can now divide just as before, only some small part of the work is distributed across the brain and the silicon circuit: a genuinely mental process (division) is supported by a hybrid biotechnological system...If you imagine a case,

identical to Diva's, but in which the restored (or even some novel) functionality is provided ... by a portable device communicating with the brain by wireless, it becomes apparent that actual wiring is not important. If you next gently alter the details so that the device communicates with Diva's brain through Diva's sense organs (piggybacking on existing sensory mechanisms as cheap way stations to the brain) you end up with what David Chalmers and I dubbed 'extended minds'.

In this scenario, it would seem the embedded theorist must insist that Diva's restored arithmetic abilities merely 'depend' on or are 'coupled with' the external silicon circuit (as opposed to being, in part, realised or instantiated by it).

But we can take this thought experiment further: Imagine Diva's brain continues to undergo damage in all the areas responsible for cognition, and we continue to restore her cognitive abilities by expanding and improving the external silicon circuit so that it becomes responsible for Diva's capacities for attention, memory, reasoning, and so on – all the way until the external circuit comes to carry out every cognitive function of her brain (with two-way wireless communication with her senses and peripheral nervous system). Experientially, Diva is not aware of anything different about what it's like for her to perform cognitive tasks.

This would be an instance (an admittedly purely hypothetical one) where an external prop not only plays an important explanatory role in cognition, but seems to instantiate the entirety of cognition. And yet, for embedded theorists, it seems all they can say is that Diva's cognition now fully depends on the external circuit. This is strange: the term "depends" here seems to function in the way "my cognition *depends* on my brain processes" – which is to say, we have shifted from defining 'depends' as "relies on", to "depends" as "relies on because it is also formed by it" (as in the case with my cognition's "dependence" on my brain). In fact, if we interpret the embedded theory strictly as a claim that only neurons can constitute cognitive processes, it seems to suggest Diva has lost the capacity for cognition altogether since her neurons for cognition have been damaged – the embedded theory cannot make sense of her seemingly restored cognitive abilities.

More importantly, the embedded theorist must find a non-arbitrary answer to this question: what would be the significance of placing this external circuit inside Diva's skull? What is it about that change in location that transforms her cognition from "depending" on the external circuit (or even, not existing if we interpret the theory uncharitably) to becoming constituted by it?

If the answer is nothing changes – we have merely moved the environment into her body – then it becomes clear the embedded theorist is using the term "depend" in the sense of "my cognition depends on my brain". What this suggests is the embedded theory does not have the resources

to answer this question without some semantic sleight of hand that redefines what “dependence” means whenever it suits the theory. EMT, on the other hand, does not have to struggle with this question: The external circuit behind Diva’s restored cognitive abilities *is* what constitutes her cognition, regardless of whether it is kept outside or inside her skull.

It is worth emphasizing that the hypothetical nature of Diva’s story is not relevant. As Clark (2009) points out, this is in principle possible, as evidenced by a Californian spiny lobster that had an “electronic neuron” successfully integrated with neural framework so that its signalling rhythm fell into place with the other cells to make it possible for the lobster to digest food. No doubt human cognition is more complex than lobster digestion, but there is no reason to think it is not in principle possible to mimic its neural underpinnings electronically in an external device, integrating it back with other neural and sensory processes wirelessly. Indeed, this would not be very different from recent experiments where subjects communicate “telepathically” while miles apart through so-called brain-to-brain communication (Grau et al., 2014; Tressoldi et al., 2014), except in this case one of those brains would be an electronic circuit.

What this shows is that a more conservative reliance on the embedded theory can reflect a kind of dogmatism that can only be upheld through an inconsistent use of the notion of “dependence”. To that degree, EMT offers a more coherent way of understanding cognitive processes. What this means is, when discussing changes to a person’s psychology, this may include a far larger domain than we might have presumed.

3. The implication of the extended mind thesis for cognitive enhancement

If EMT is true, what can this tell us about our attitudes to changes in our psychology? The implication of EMT is that we regularly undergo such changes through the use of external cognitive props – to the extent that those changes are advantageous, we have no qualms, in general, about cognitive enhancement. That is to say, while we can and ought to question how we develop technologies and their impact on the environment and our lives, there is little that is ethically contentious about upgrading, say, one’s smart phone or computer.² We don’t see such upgrades as raising the same moral issues as upgrading aspects of one’s internal mind. And

² Of course we may have ethically motivated environmental and/or anti-consumerist reasons to oppose such upgrades; however, here I am only referring to ethical qualms about the advantages of a better phone or computer.

yet, according to EMT, such upgrades (or downgrades) can literally constitute upgrades (or downgrades) to our minds, at least in certain circumstances.

If we find such external changes generally acceptable, shouldn't it follow that upgrading the internal mind, through intervening in the brain, is also acceptable – in principle at least? Clearly intervening in the brain might pose greater risks and have broader or less predictable consequences, but it seems to follow that, in principle, there is no distinction between internal and external interventions. What matters is only the costs and benefits between these two options. Or, as Levy (2011) puts it:

Much of the heat and the hype surrounding neuroscientific technologies stems from the perception that they offer (or threaten) opportunities genuinely unprecedented in human experience. But if the mind is not confined within the skull...[then] intervening in the mind is ubiquitous. It becomes difficult to defend the idea that there is a difference in principle between interventions which work by altering a person's environment and that work directly on her brain, insofar as the effect on cognition is the same; the mere fact that an intervention targets the brain directly no longer seems relevant.

This observation leads Levy to develop an Ethical Parity Principle (EPP). The parity principle introduced by Clark and Chalmers entails the possibility of cognitive processes being constituted by both internal and external vehicles. EPP takes this conclusion further, and suggests that, therefore, there is also a moral equivalence between interventions in internal and external vehicles that constitute parts of our cognitive processes.

Levy (2011) formulates EPP in this way: Since the mind extends into the external environment, alterations of external props used for thinking are (*ceteris paribus*) ethically on a par with alterations of the brain.

The use of *ceteris paribus*, which is to say, all other things being equal, means Levy is talking only in principle. As stated, we may have all sorts of reasons to think an alteration to the environment is better than to a brain (and vice versa). This may be because of the costs, risks, and practicalities involved, but it is not because there is anything about alterations in the brain that, in and of itself, is morally significant, or rather, more morally significant than alterations in external props that come to constitute parts of our cognitive processes.³

³ Of course, the history of early attempts to intervene in the brain through leucotomy and lobotomy may have coloured our perceptions of internal interventions – needless to say, such interventions horrify us specifically because of how disabling (as opposed to enhancing) they were.

An example may help illustrate this: Clark (2003) notes the case of Alzheimer's sufferers who successfully live alone in city apartments, even though psychological tests suggest these individuals really should not be able to cope with everyday demands. The answer to this riddle turned out to be that these Alzheimer's sufferers had basically turned their homes into an extension of their minds: cognitive props and aids were pervasive in their apartments, with labels and pictures on doors, memory books like Otto's to record new events and plans, photos of friends and family with their names and relation to them, with crucial daily items like cooking utensils kept in plain view.

In such a case, EPP suggests that, say, forcefully relocating such individuals is ethically on par with erasing a cognitively healthy person's long-term memories – all else being equal.

However, all else being equal is starting to look suspicious at this stage. If it turns out all else is rarely equal in regards to alterations to the brain versus the environment, then this EPP may not be particularly useful. That, to an extent, is the argument DeMarco and Ford (2014) make. They critique EPP and its attempt to free neuroethicists from an alleged prejudice against internal enhancement. They point to some differences between the brain and environment that are not necessarily "in-principle" differences, but that are nevertheless important moral differences that imply all else is rarely equal.

One such example is that external props are more replaceable than brain parts. Levy (2007) himself points out that "Otto's notebook is far more easily replaced than his hippocampus." But the fact that external props are more fungible seems to be a morally relevant difference – perhaps not in principle, but certainly in fact. It is a qualitative difference that seems inherent to the brain in contrast to external props.

DeMarco and Ford (2014) also point out the fuller integration of the internal mind as another qualitative difference. In the case of memory, for instance, internal memory is in a dynamic state: new memories have implications for older ones. For instance, if Inga were to learn the museum has burnt down, that would have implications for what she believes about its location. Her belief about its location becomes a belief about where the museum used to be. Not so for Otto: new entries do not make automatic changes to older entries.

A third disanalogy DeMarco and Ford (2014) point to is consciousness. Even if in some cases external cognitive processes come to literally constitute part of the mind, the brain's role as a necessary part of this extended mind indicates an important moral distinction to be made between what's internal and external. Altering the brain may impact consciousness in a way that

alterations to external props cannot: If, unbeknownst to me, my smartphone were being crushed at this moment, my consciousness would remain as is. Not so if it were my brain being crushed. In that sense, it is difficult to see how an alteration to the external environment may be on par to an internal one in terms of consciousness.⁴

It is not clear the clause *ceteris paribus* covers these seemingly fundamental differences. I say seemingly because, of course, it is possible to envision future technologies that nullify such differences. While replacing a hippocampus is currently impossible, it may not be so forever. It is conceivable that in the future such a replacement would be on par with replacing a gadget. Similarly, while Otto's notebook lacks the kind of informational integration that biological memory has, smart journals can organize information in a more similar way to biological memory. As for consciousness, it is possible to imagine future devices that are deeply integrated into our mental models of the world in ways that change how we perceive it (Danaher, 2015).

The point here is that these differences, while important and pose a problem for the ethical parity principle, do not reflect necessarily inherent differences between the internal and external mind. They may become immaterial differences in the future. Until then, however, it seems EPP's implications will be limited: Otto's notebook is replaceable and its loss does not impact his self-consciousness or his understanding of himself – burning it would not be quite like erasing a cognitively healthy persons' long-term memories.

However, it is not clear that these differences suggest what DeMarco and Ford (2014) claim they do. To be clear, DeMarco and Ford want to suggest that favouring external enhancements is not merely based on prejudice but on the salient moral differences they raise. However, as Danaher (2015) points out, these actually give us a reason to favour internal enhancements (that is, enhancements directly to the brain).

On the one hand, these differences suggest there are greater moral risks associated with interventions targeting the brain. On the other, they show that interventions to the brain that are safe and have well understood consequences may be preferable to external alterations. If the argument is that the brain, unlike external props, is not replaceable and without it consciousness is not possible, then perhaps it is the brain we should prioritize enhancing (when that is an option). Having consciousness is an ethically salient property that is central to

⁴ These issues facing EPP highlight an instance of the problem noted in chapter one regarding the apparent conflation of abstract and concrete ethics. The EPP is a great example of abstract implications arising from a conceptual analysis of cognition, where converting those implications to concrete, real-world questions turns out to be of somewhat limited relevance.

discussing what it means to live a good life. This seems to suggest it is that which we should prioritize enhancing.

Consider this case: Otto could improve access to his memories in two ways. One is to take a daily pill that will kick start a process of neurogenesis in his hippocampus and restore his memory formation and recall abilities. It is both safe and affordable. The other option is an improved version of his trusty notebook: it keeps an online copy of the information in it in case it is lost or damaged, and entries that have implications for previous entries are automatically updated. Again, not all is equal here: the notebook is replaceable while his hippocampus is not, and the use of his notebook has a different, less integrative phenomenological experience to it compared to accessing or updating his brain-based memories. And yet, those differences give Otto a reason to favour taking the drug: being able to access his memories more directly through his brain and to experience the phenomenological qualities associated with doing that is advantageous to him in a way that relying on the notebook is not. Similarly, if there were a safe pill that, say, protected Otto's previously healthy hippocampus from degenerating, we ought to favour that over, say, him filling his notebook early on so that it is ready and suitable for use after his hippocampus degenerates. Again, this is precisely because the irreplaceability and phenomenological qualities of having a hippocampus (compared to a back-up device) is a morally relevant reason for us to favour the internal intervention.

In other words, what DeMarco and Ford's argument may actually imply is that, if met by two ways to change a person's psychology, one that acts by modifying the external environment and the other by modifying the brain, so long as both are similarly safe, affordable and have well understood consequences, we should be partial to the latter intervention. Of course, how we weigh up those factors will depend on the specifics of the case: it may be that even if the internal intervention is not as safe as the external one, its other benefits outweigh the risks posed by its safety. Hence, the "relevant set of circumstances" of a person will play a significant role in such deliberations, but the more crucial point here is merely to illustrate that the fact that not all is equal does not mean we should be partial to external interventions – it may be quite the opposite.

4. Conclusion

To summarize, let us return to the relevance of this for the welfarist account of enhancement: if a person's psychology is not limited to their brain, then some changes to the environment that form part of a person's extended mind can be enhancements if they tend to improve well-being. Similarly, such environmental changes that tend to reduce well-being would be disabling. Hence, if we remove someone with Alzheimer's from an environment that they use as part of their extended mind to compensate for (or restore) their loss of neural functioning, then doing so may be disabling for them in the welfarist sense. Nevertheless, it seems relevant to acknowledge that the internally realised aspects of our psychology are (at the moment) more vulnerable and play an integral role to our consciousness. However, I have argued that, in some cases at least, this may in fact give us a reason to favour enhancing those internal mechanisms rather than external ones.

Linking this to the broader thesis, we can see its importance in helping us think about enhancement. Enhancement, as it pertains to our psychologies at least, is not limited to internal changes to our bodies or brains. Wariness against certain enhancements cannot be justified purely because they are internal – more needs to be said about what it is that makes the fact that they are internal relevant. This conclusion will be important for practical ethicists and policy makers: regulating interventions that work internally differently than ones that work externally *merely* because they are internal would reflect an unjustified bias.

In the meantime, a brief recap of where we stand is in order. I have at this point arguably examined the implication of each component making up the definition of enhancement used by the welfarist account. We have examined the implications of the focus on “any changes”, of the overlapping consensus on what constitutes a good life, of the “relevant set of circumstances” and how enhancement relates to disability, and now, of what is entailed by changes to a person's psychology. Together, these reflect quite a broad account of enhancement: whether a change is restorative of health or not, improving or diminishing of functioning, science-based or conventional, impacting the brain directly or external cognitive processes, all can count as enhancements so long as they tend to improve one's chances of leading a good life – something that is entirely determined by one's particular circumstances.

As such, the welfarist account appears to be a general account of enhancement: whether the change is reflected in one's genes, body, mood, cognition, or in their longevity, the account seems able capture and make sense of it. However, a crucial item is missing from the above

forms of enhancement: changes to our psychologies that impact our moral dispositions and behaviours. Can the welfarist account capture this form of enhancement too? That is the topic of the next chapter.

Chapter 4

Welfarist Moral Enhancement

One area of enhancement that has generated a massive, often heated, debate is moral enhancement: interventions often characterized as leaving an individual with more moral motives or behaviours than they otherwise would have had (e.g. Douglas, 2013). However, as I have previously noted, the welfarist account of enhancement, which roots its definition of the term in what is advantageous for a person, seems to conflict with this: surely becoming more moral can, at least sometimes, not be advantageous for a person, such as when doing the right thing entails a significant sacrifice on one's part. If so, the welfarist account might deem interventions that improve moral motives or behaviours, but that come at a personal cost, as disabling. This seems troubling for the welfarist about enhancement, raising similar concerns as the mislabelling objection we encountered in chapter two. With that in mind, this chapter attempts to answer this question: can the welfarist account make sense of moral enhancement? It is a question that has not been raised by the original formulators of the account, or the critics of the account we encountered in chapter two.

The chapter unfolds in four parts. First, I more fully articulate the seeming conflict between the welfarist account and moral enhancement. I argue that while we may wish to distinguish welfarist enhancement from moral enhancement, those two are likely to conflict far less often than might be assumed. In fact, I argue the welfarist account has an important role to play in helping ethicists and policy makers assess different forms of technologies that impact our moral motives and behaviours. Second, I argue that the welfarist account can also help with better understanding some of the current disagreements about the ethics of moral enhancement, which tend to centre on concerns about how moral enhancers might disadvantage those who make use of them. As such, I argue for an approach to moral enhancement that is indirect, focusing on the (welfarist) enhancement of certain aspects of cognition that lend themselves to improved moral reasoning and motivation. To support this approach, in the third section I examine the evidence for the relationship between specific cognitive capacities and moral behaviour. Finally, I explore the notion of moral motivation enhancement and how it might relate to certain cognitive capacities. Altogether, my goal for this chapter is to highlight a limitation of the welfarist account – it lacks the scope to capture all instances of moral

enhancement – but also a positive feature of the account: it has a role to play in assessing moral enhancers and it may offer a less contentious, indirect approach to moral enhancement.

1. The welfarist account of moral enhancement

How might the welfarist account define moral enhancement? To the extent our moral motives and dispositions are part of, or result from, our psychology, these would seem to be covered under the broad definition of enhancement as a well-being improving change to our psychology. More specifically, we might articulate it as: Any change to the moral motives of a person which increases their chances of leading a good life in the relevant set of circumstances. We can change moral motives with moral capacities or moral behaviours if we wish. But, whichever way we articulate it, we can already see the implausibility of the above characterization of moral enhancement.

This implausibility stems from the fact that, according to such a definition, if one becomes more moral but leads a worse life as a result, it seems they have become disabled. Such a moral change is disadvantageous for them. In fact, for something to count as a moral enhancement in the welfarist sense, it must always be advantageous to the person who has undergone the alteration. This seems a strange way of articulating the nature or value of moral enhancement and arguably limits the scope of the welfarist account. Indeed, some of the architects of the welfarist account seem to conclude as much, quietly disregarding it in favour of a functionalist approach to enhancement where improving or augmenting our moral capacities is what constitutes moral enhancement (Earp, Douglas, & Savulescu, 2017; Persson & Savulescu, 2008, 2012).

On the other hand, it is not entirely clear what moral enhancement entails exactly. As a goal, it appears a common feature of all societies, being pursued through moral education, engagement in moral literature and religion, and perhaps also certain meditative practices. More recently, it has been explored as a direct prospect, with studies investigating the use of hormones such as oxytocin or non-invasive brain stimulation to potentially increase pro-social attitudes (Nave, Camerer, & McCullough, 2015; Young, Camprodon, Hauser, Pascual-Leone, & Saxe, 2010).

Aside from the technique used, moral enhancement itself has been characterized in ways ranging from merely feeling more empathic concern, to increasing one's sense of personal responsibility, to having a heightened respect for global fairness (Shook, 2012). For instance,

Tom Douglas (2013) characterizes moral enhancement as an intervention that will expectably leave an individual with more moral motives or behaviours than they would otherwise have had. On the other hand, Harris (2014) sees it as enhancing our ability to think ethically rather than a direct modulation of our motives or emotions. Alternatively, Ingmar Persson and Julian Savulescu (2008) are concerned with the “moral character of humanity”, including its underlying genetic basis. Halley Faust (2008) considers the hypothetical possibility of using genetic screening to select for embryos disposed to a higher level of morality. On this approach, some may argue that no resulting person could claim to be morally enhanced – at least on an account like Douglas’s – since they would not have existed if not for already having those genetic moral dispositions. In that respect, how we define moral enhancement may depend on whether we are considering individuals or humanity more generally (Raus, et al., 2014).

Moral enhancement is further complicated by the fact that moral judgements are not identical across the world. Unlike enhancing a cognitive capacity like attention – a property that can be broadly understood independently of social context – individuals and cultures can understand morality differently (Graham et al., 2016). But morality is also context dependent in a more fundamental way, making it possible for an enhanced moral disposition (e.g. being tolerant) to lead to potentially morally deficient behaviour (e.g. tolerating injustice).

Within this quagmire, the welfarist account of enhancement adds a further difficult question: what is the relationship between what is advantageous to a person and what is moral for them to do? It is a conflicted relationship, often expressed as prudential versus moral value, or personal well-being versus morality more generally (R. Chang, 2004). That conflict stems from the fact that, at least sometimes, being moral can reduce one’s well-being, while pursuing one’s self-interest can often be immoral.

Michael King (2016) conveys such a possibility, using utilitarianism as a reference point:

One could imagine an alteration to biology or psychology that made the individual much more willing to act on utilitarian demands to sacrifice their own well-being for others until doing so would fail to promote aggregate utility, which would make overall loss of prudential value for the enhanced individual not only possible, expectable. Since they do not expectably increase well-being overall for the individual, these putative moral enhancements would not be enhancements according to the [welfarist] account, and the account does not provide any reason to choose them.

King takes this to its logical conclusion: the welfarist account would not just reject this as moral enhancement – it would deem the alteration a disability or disadvantage. However, we need not assume utilitarianism for this point to stand: we merely need to imagine actions that create

moral value (whatever it is) that conflict with prudential value. What such examples may suggest is that the welfarist account lacks the conceptual resources to successfully refer to changes to an individual's dispositions that make them morally better, but that leave them expectably worse off. That is, its scope is not sufficiently general to capture moral enhancements.

This, I think, is broadly correct. An intuitively appealing way to make sense of why people would seek to change their bodies or minds is because of a belief that doing so will improve their well-being. Being smarter, stronger or longer-lived hold clear prudential benefits. Yet being more moral seems different. True, many traits associated with being moral can makes our lives go better: leading a more tolerant, patient, and understanding life would seem to be a life with less frustration and hatred in it, and possibly more friendship and solidarity, and in that sense inducing those traits in a person who lacks them might be advantageous for them. Yet it is not clear that other traits associated with being moral, such as being compassionate or selfless, are advantageous in the prudential sense: these may entail leading a more difficult or burdensome life. Because of that, it seems plausible to assume most people would at least hesitate to undergo some intervention that might mean we will wish to suddenly, say, live more frugally, or donate much of our wealth, or go on vacation less often, or spend more time volunteering in night shelters. We would hesitate precisely because it's not clear that such a change to our lives would be advantageous from a prudential perspective. And so, if we were to nevertheless undergo such an intervention, it is likely to be motivated by something other than the promotion of our own well-being.¹ In that way, moral enhancement does seem fundamentally different from welfarist enhancements such as increased strength or intelligence.

Because of that, we may wish to distinguish between welfarist enhancement (which is self-regarding by definition), and moral enhancement (which is not necessarily self-regarding) as two separate domains of enhancement. Here, welfarist enhancement includes all forms of enhancement to oneself (physical, aesthetic, genetic, cognitive, mood-based), *except* moral enhancements. Moral enhancement is different in the sense that it can be best understood in a functionalist, augmentative sense.

However, to conclude from this that the welfarist account is simply not relevant to the discussion on moral enhancement would be mistaken. For the remainder of this section, I will argue that much of the seeming tension between welfarist and moral enhancement boils down to articulating it in abstract terms. Keeping in mind what I noted in chapter one on the distinction

¹ This is not to ignore the fact that there is nevertheless a dispute in theories of well-being as to whether moral virtue is good for an agent (Hooker, 1996).

between abstract and concrete ethics (and the role those can play in our conceptual analyses), I will argue here that once we attempt to ground this tension in concrete examples, we will see that the welfarist account has an important role to play in capturing what is often most salient about interventions that change the moral motives or behaviours of a person (I will refer to such interventions as “moral technologies” if they are biomedical in nature) . To do that, I present four thought experiments where someone’s moral motives or behaviours are seemingly improved, but where they personally become worse off. I argue the welfarist account’s referral to the interventions in these concrete scenarios as disabling is correct, and reflects an important dimension to our assessments of moral technologies that would be lacking in functionalist accounts of moral enhancement such as Douglas’s (2013). In line with my methodological approach to this thesis, note also that I am not speculating about the possibility or likelihood of these hypothetical moral technologies ever being developed – instead, I refer to them merely to draw observations about the scope of the welfarist account.

The clearest example of improved moral motives or behaviours that leave someone worse off is sacrificing one’s life to save another’s. Here, the cost to the individual is overwhelming: not only might the action be costly in terms of the pain or fear entailed by the process of dying, but it is supremely costly in that they lose the capacity for well-being altogether.² And yet, such an act would ordinarily be understood as moral, if not heroically so.

Based on this, suppose a drug is invented, call it the Jesus drug: you take it and suddenly feel highly motivated to find a way to sacrifice your life to save others. You track down terminal patients who need a heart, liver, and lungs and with whom you match. You notify the relevant authorities, and kill yourself in the appropriate place, saving three others.

The Jesus drug, as is quite clear, is highly disadvantageous for the user. True, it augments a particular moral motive (sacrificing yourself for others), but that is clearly not the salient feature about it: the salient feature is that it makes its user end their lives very quickly and reliably. Calling it a moral enhancer merely because it augments a certain moral motive seems to be missing the more central point about what it does. The welfarist account, on the other hand, which would label it a disabling drug, captures what is arguably its most relevant effect. In other words, the welfarist account is able to reflect what is horrifying about this moral technology in a way that a functionalist account of moral enhancement cannot.

² Unless one believes in an afterlife – however, I am referring to death here as the cessation of an individual’s consciousness, not its transference to some other alleged realm.

Of course, perhaps being moral requires reaching decisions more deliberately – the Jesus drug just creates one overriding desire to sacrifice one’s life, and in that sense some may argue it does not qualify as a moral enhancer. If so, then this is not an instance of being more moral that is disadvantageous, and thus does not reflect a tension between welfarist enhancement and moral enhancement.

Let us consider a slightly less extreme case that captures King’s (2016) abstract example of a super-utilitarian who maximizes aggregate utility but at the overall loss of prudential value to themselves. Suppose an individual, call her Beneficent, hears about a new drug, *Altruix*, that will improve her moral motives and behaviours. Wanting to be a morally better person, Beneficent starts taking it and soon after notices that she wants to be kinder, more understanding and generous, and so begins behaving in those ways more and more often. These changes improve the lives of people she comes into contact with (as she is a much more caring and helpful person now), and this in turn improves her own life too: there appears to be more love in her life and she feels she contributes to others in a way that is meaningful. A sense of purpose ensues that overwhelms any of the sacrifices and efforts entailed by being more caring and giving. Up until now, her moral actions have been prudentially advantageous to her. In other words, *Altruix* appears to be a moral enhancer and a welfarist enhancer.

However, as Beneficent continues to take *Altruix*, things get out of hand: she begins donating more and more of her time and income to others, losing touch with friends and family. Most of her furniture she sells and donates the money. Eventually, she donates a kidney to a stranger. In fact, she finds that over the coming months she keeps herself only as healthy and wealthy as is necessary to maximize her ability to help others. What’s more, these further developments are now severely diminishing her well-being: she feels constantly exhausted and in pain, crying herself to sleep every night at the thought of others’ suffering. More than anything, she now feels deeply alone in the world as few others share her concerns in the same way. Nevertheless, there is no doubt her actions continue to improve the lives of people around her (and all around the world where she has sent her donations).

Clearly, Beneficent, while she continues to create moral value in the world through her actions, has suffered a significant diminishment to her well-being.³ From a prudential perspective, she was better off before taking *Altruix*. With that in mind, note that the welfarist account seems very capable of capturing this fact about Beneficent’s life: the use of *Altruix* has had a disabling

³ We can envision a deontological version of this scenario. Here, taking *Altruix* motivates her to live by the categorical imperative in ways that similarly diminish her well-being.

effect on her. It has proved to be a disadvantageous change to her psychology. Is there anything particularly lacking about this conclusion? Indeed, this seems to be the more salient fact about what *Altruix* does to its users. The welfarist account does not need to deny *Altruix* is a moral technology, and we may wish to call it a moral enhancer in the functionalist sense – the drug does indeed dispose people to behave in more moral ways. Yet merely calling it a moral enhancer seems to be missing something fundamental about what the drug does: it eventually destroys the lives of people who take it. From the perspective of deciding to make such a drug and provide it to people like Beneficent, surely that is the salient feature worth highlighting about this drug: what people need to know about it is that, first and foremost, it will destroy their lives, not improve their moral dispositions. The fact that it is a “moral enhancer” in the functionalist sense is relatively irrelevant. In that way, the welfarist account plays an important role in capturing what this intervention does.

Another scenario can further convey this point: Imagine Pido, a paedophile, is regularly injected with a drug against his will that makes him lose all urges to abuse children. He now has better (or at least less bad) motives towards them. However, this drug acts on a group of neurotransmitters central to mood regulation: Pido finds that while he is freed from those urges and from the risks of incarceration and/or further ostracisation, he now suffers from deep depression: he does not eat or bathe, and seems to have simply lost the will to live. His improved moral motives have come at a greater prudential cost to himself.

Again, we seem to have an instance of what is arguably a moral enhancer that is disabling from a welfarist perspective. Note that a society could decide it has good reasons to disable Pido in this way despite its effects on him, and the welfarist account is not incompatible with such a conclusion. That is, this may be a case where severely disabling someone is the right action, all things considered. But note also how the welfarist account highlights what should arguably be occupying the bulk of our ethical deliberations in such a case: should we destroy a person’s life in order to reduce the chances of him harming some children in the future? A functionalist approach to moral enhancement is unable to be nuanced in this way. It does nothing to bring out what is at stake exactly nor highlight the difference between changes to moral motives that ruin a life or improve it. In that way, the welfarist account, even though it limits its scope to self-regarding interventions, plays an important role here, creating the kind of framework conducive to deliberating clearly about what is at stake exactly: the chances of a person’s leading a good life and its relation to broader values (be it justice, reducing overall harm, and so on).

Perhaps, however, these are all easy cases. What about a moral alteration to a person that only leaves them very slightly worse off? Is the loss in prudential value in this case still the salient feature of such an intervention? The first thing to note here is that it is not easy to articulate what such a case entails. Suppose Grumpy is an elderly man who is a little racist and does not like to share his food or help his neighbours. He's given a course of *Altruix Light* and becomes ever so slightly more tolerable: he starts waving rather than frowning at the Filipino couple next door and his wife actually enjoys his company now. He himself starts to appreciate his wife and local community more. Despite this, we can imagine he has become slightly worse off than before: perhaps he spends less time alone now and therefore works on his carpentry less often, and so ends up making less beautiful things in his life. Without him realising it, the drug has also been the cause of regular headaches that now diminish his ability to enjoy watching television or reading.

The problem with cases like Grumpy's is that they force us to adopt a clear conception of well-being in order to evaluate them. In the previous cases, some overlapping consensus about what well-being entails was sufficient to conclude the interventions in them were clearly well-being diminishing for the individuals involved. Here, it seems like we might get different answers about how well Grumpy's life is going now depending on whether we are hedonists, desire satisfactionists, or list theorists (and for the list theorist, the contents of that list and how we weigh up its constituents is crucial).

Nevertheless, the fact that the welfarist account requires us to become clearer about what well-being is in such cases (and how we ought to weigh it against promoting other values) is a positive feature of the account that is relevant to moral enhancement: surely whether *Altruix Light* should be made available to only mildly morally deficient individuals ought to depend on how it impacts the well-being of people who use it. If it will reliably make users' lives go worse while only marginally improving moral motives and behaviours, it may be hard to justify its use. And the reason for that is precisely because of its impact on prudential value. And so, again, the welfarist account is keeping the evaluative focus on the more relevant issue here. Functionalist approaches to moral enhancement are not capable of this.

Of course, an alternative here is, after some further thought, we conclude that, overall, the negative and positive effects of *Altruix Light* on Grumpy's well-being cancel each other out: it has had no effect on his well-being overall. In such a case, the welfarist account would not label *Altruix Light* as enhancing or disabling. It would be a moral technology that simply has no noteworthy effect on the user's well-being. Here, it is precisely such cases that would highlight

the welfarist account's lack of conceptual resources to capture everything there is to say about moral enhancements: when a moral enhancer has no implications for a user's well-being, the account is silent, giving us further reason to distinguish welfarist enhancement from moral enhancement.

Nevertheless, it is worth emphasizing how difficult it is to imagine an intervention that alters a person's dispositions without either some net benefit or cost to them. Perhaps this will be the case for individuals in very specific circumstances with very specific interests or values, but it seems implausible to conceive of a moral technology used by, say, millions which consistently has a prudentially-neutral effect on their lives. Much more plausible is that it will either be disabling or enhancing for most of users.

What all this reflects is the important role the welfarist account has to play in understanding moral enhancement. It may be that distinguishing the two (i.e. welfarist and moral enhancement) is necessary in order to capture the full scope of possible enhancements, but to completely divorce the two would be mistaken: when discussing a moral enhancer, a crucial feature of it should be how it impacts the well-being of individuals who use it. In that sense, we should think in terms of advantageous moral enhancers, disadvantageous moral enhancers, and prudentially-neutral moral enhancers. This relationship between what is prudential and moral will be key in our assessment and characterisation of moral technologies, and it is only through a welfarist lens that we can make that more complete assessment. In fact, it is worth spelling out that relationship in detail within this context. That relationship can be articulated in twelve ways.

Changes to an individual's biology or psychology can leave a person:

1. Prudentially advantaged while also increasing moral value: These are enhancements that we also have non-prudential reasons to favour. An example may be vaccines. These can be advantageous for you, but for others as well, who are less likely to get sick precisely because you are less likely to get sick. These are the sorts of enhancements society is likely to want to promote.
 - a. Prudentially advantaged as well as more likely to act morally. This is a subset of the previous category. We can term these interventions welfarist moral enhancements. It is a subset because it is prudentially advantageous and also increases the chances of moral value, but does so in a specific way through the intentional behaviour of the individual. There may be many examples of welfarist moral enhancements: becoming kinder, more patient, more generous.

These are welfarist moral enhancements only to the extent they also tend to improve the well-being of the individual compared to how it otherwise would have been.

2. Prudentially advantaged while reducing moral value: These are controversial enhancements. An example of these may be enhancements that confer a primarily competitive advantage or that lead to more discrimination in a society. Some examples may include cosmetic enhancements, altering one's racial features or sex, and so on.
 - a. Prudentially advantaged but less likely to act morally. This too is a subset of the previous category to the extent it also reduces moral value, but specifically through the behaviour of the person. These would also be considered controversial enhancements. We can imagine societal or parental interventions that change someone's beliefs so that they think they are more moral without having to act morally (this would work on a hedonistic account of well-being). Alternatively, we can change someone's preferences so they have fewer or weaker moral desires and more or stronger selfish and easily fulfilled ones, thereby potentially increasing their well-being.
3. Prudentially disadvantaged while increasing moral value: These are disabling or disadvantageous but nevertheless morally valuable interventions. An example could be a cognitive enhancer that leaves the user feeling alienated from others and miserable, but nevertheless able to contribute highly morally valuable ideas to society. The welfarist account gives us a *prima facie* reason to be hesitant about such interventions, but it is compatible with an endorsement of them when other non-prudential values are taken into account.
 - a. Prudentially disadvantaged but more likely to act morally: This subset of the previous category refers to a disabling moral alteration to a person. A case in point is Pido: he no longer has urges to abuse children but is now severely depressed. As should be clear from the previous discussion, we are unlikely to find many such cases as, in general, being more moral is normally entangled with prudential value.
4. Prudentially disadvantaged and reduces moral value: a disabling and useless intervention from the moral point of view. An example might be a highly addictive and damaging drug.
 - a. Prudentially disadvantageous and also less likely to act morally: This subset is also another useless and harmful intervention.

5. Prudentially the same but increases moral value: this is neither a disabling nor enhancing intervention, but it is nevertheless a valuable one. For instance, if it turned out being smarter has an all things considered prudentially neutral effect on a person (say it improves their life prospects but also makes them prone to depression in ways that cancel each other out), it may nevertheless increase moral value if that intelligence is used towards a valuable end.
 - a. Prudentially the same but more likely to act morally: I have suggested it may be very difficult to think up of such an intervention, but to the extent one hypothetically exists, it is neither enhancing or disabling for the individual who undergoes it.
6. Prudentially the same but decreases moral value: an intervention that would harm others but will not be sought by anyone as it provides no advantage.
 - a. Prudentially the same but less likely to act morally: same as above, except the harm to others would come directly from the actions of the person who undergoes such an intervention.

The scenarios in the four thought experiments above deal with 3a: a disabling but morally improving change to a person. It is true, as King (2016) argues, the welfarist account provides no (initial) reason to choose these; in fact, it gives us a reason to be suspicious of them. But, as I have argued, that is surely a strength of the account: any intervention that will expectably make someone worse off *should* give us pause. Moreover, if the labels we use ought to capture what is most salient about something, the welfarist account does so by labelling such interventions as disabling for the user. As we have seen in the case of the Jesus drug, Beneficent and Pido, it is clear those are the more salient features of the interventions. This is not to deny there might be cases where it is unclear what is more salient, as in the case of Grumpy. In such a case, the welfarist account at least brings our attention to what seems most relevant to deciding whether to make or distribute such an intervention: the impact on the individual who takes Altruix Light.

On the other hand, if overriding reasons can be identified to make a person worse off, then we may have good non-prudential reasons to make use of a disabling moral enhancer. Nevertheless, this way of thinking about moral enhancement reflects why we ought to, in general, favour moral technologies that fall under 1a – that is, welfarist moral enhancements. These are examples of ‘win-win’ moral technologies, where the two “wins” refer to prudential and moral value.

In conclusion, the welfarist about enhancement can concede that its referral to interventions as moral enhancements only when they have a self-regarding benefit seems at odds with the nature or value of moral technologies. Because of that, we may wish to distinguish welfarist from moral enhancement, whereby the latter are understood as a functionalist enhancement, with the former being used to unpack their prudential value. As such, even with this distinction, the welfarist account still has an important role to play in guiding our thinking about moral technologies. In that sense, the framework that the account sets up would be highly relevant to deliberations about the ethics of different moral technologies or interventions.

2. Cognitive enhancement as indirect moral enhancement

In this section, I argue for another way the welfarist account is important to the discussion on moral enhancement; namely, how it can help clarify some of the disagreements in associated literature.

An ongoing dispute about moral enhancement is whether interventions that directly act on people's attitudes and motivations are permissible (Agar, 2014; DeGrazia, 2014; Douglas, 2014; Harris, 2011, 2014, 2016, Persson & Savulescu, 2016a, 2016b). John Harris (2011) argues interventions that would tinker with people's attitudes and emotions so as to increase the probability they will act in ways deemed ethical would rob them of important aspects of their autonomy – namely, their “freedom to fall”, which I take to mean our ability to choose to do wrong. In that sense, Harris is concerned such forms of biomedical moral enhancement would come at a grave prudential cost to the individual. Schaefer (2015) raises a variation of this concern along the lines of the loss of individuality, which he defines in Millian terms as our capacity to hold beliefs and motives that are our own – that is, ones we have cultivated or developed ourselves rather than merely conformed to. Lacking such individuality can reduce us or bring us closer to being passive, mechanistic beings vulnerable to the instrumental control of others. In that way, if people's moral dispositions are directly modulated through an enhancer, this may reduce us to just that: mechanistic beings incapable of moral disagreement. Again, this concern, to the extent that it is valid, is a concern about our capacity to lead good lives.

While I do not intend to weigh in on whether such interventions would be prudentially costly in that way (that is, that they might rob people of their autonomy or individuality), it seems useful for ethicists and policy maker to be able to clarify what those concerns are in welfarist terms: if

Harris and Schaefer are correct, then such interventions may be severely disabling to users, even if they do improve moral dispositions.

In response, their critics may argue that reducing the chances of great harm in the world may be worth the prudential cost of slightly diminishing individuals' autonomy (Persson and Savulescu 2016a), while others may argue such interventions would not in fact reduce autonomy in any relevant way since we are not the authors of our moral dispositions anyway, and therefore moral enhancement would not be an added prudential cost (DeGrazia, 2014).

Setting out the dispute in these terms helps to clarify what these disagreements are about: moral value versus prudential costs.⁴ What is clear is that most agree changes to our moral dispositions should, ideally, not come at a prudential cost to those who undergo such changes. That is, it is preferable such interventions are instances of welfarist moral enhancement: they increase moral acts while also being prudentially advantageous to the person undergoing the change.

To that end, the remainder of this chapter explores one avenue for approaching welfarist moral enhancement. My motivation for this is to make the preceding discussion less abstract and more substantive. It is also a novel avenue through which to explore the broader implications of the welfarist account. In this context, that avenue is indirect, using cognitive enhancement, and specifically, enhancing one's ability to think ethically. Such an indirect approach to moral enhancement has been argued as the likely 'most reliable' means (Earp, Douglas, Savulescu 2017)⁵ and the least likely method to infringe on autonomy (Harris, 2011, Schaefer 2015).

John Harris (2014) pushes for this approach as such:

I would recommend that the most effective and ethical moral enhancer so far available is learning to subject emotional reactions to the scrutiny of reason. In short, I take moral enhancement to involve enhancing our ability to think ethically (cognitive enhancement), not manipulating the probability of some reacting in ways that others deem ethical.

DeGrazia (2014) also agrees with this approach: moral judgments – or “insights”, as he terms them – are “highly cognitive and may come about through cognitive enhancement. Knowing what is right, after all, is a kind of knowledge.” Knowledge, in turn, tends to be held as

⁴ That, at least, is one dimension of the debate. Much of the discussion also revolves around a suspicion that moral enhancements would backfire – that is, they would in fact fail to actually make us more moral (Agar, 2014; John Harris, 2011).

⁵ For instance, they argue that the most promising scenario for permissible moral [neuro]enhancers entails allowing the user to consciously reflect and critically engage with whatever “moral insights” that might be facilitated by the user of a moral enhancer.

prudentially valuable, either instrumentally as a means to pleasure or a greater ability to fulfil one's desires, or as a value in itself on objective accounts of well-being.

Schaefer (2015) similarly favours this approach and elaborates on the relationship between cognitive ability and moral dispositions. For instance, improved logical and conceptual understanding, along with better reasoning abilities more generally, can help us spot contradictions in our beliefs, comprehend the implications of a moral idea, and perhaps even reduce the distorting effects our biases have on ethical deliberation.

However, while these conclusions have their appeal, they lack any empirical support. Is there any evidence that greater cognitive ability can entail better moral motives and behaviours? Answering this question seems crucial to the prospect of a welfarist approach to moral enhancement and, in line with my methodological guidelines, exploring that evidence seems crucial to reaching any such conclusion.

In general, it would seem individuals often use intelligence to manipulate information in ways that may help rationalize and justify (to themselves as well as others) all sorts of classically evil acts (Ariely, 2012; West, Meserve, & Stanovich, 2012). This should make us suspicious about any claims that more intelligence might make us more moral. Hence, in the next section, I present evidence that, on average and all else being equal, people with more sophisticated reasoning abilities are more cooperative, have larger moral circles, and are less sympathetic to violence. More broadly, I suggest a society of better-reasoning individuals would likely adopt policies less conducive to violence. I also argue moral motivation can be approached cognitively rather than through the direct manipulation of emotions, as others have suggested.

First, let me be clearer about what I will argue here. I posit there are at least four reasons for psychologically normal people to make errors in their moral judgements and acts, and that these reasons or some combination of them are at the root of most acts that most reasonable people would agree are immoral. I then argue these reasons all relate to brain states linked to cognitive abilities, and that therefore moral behaviour, in principle, can be improved through cognitive enhancement.

Here are four possible examples behind moral errors:

- 1) Epistemic factors: These may relate to being misinformed, whereby having false factual beliefs leads to immoral beliefs (e.g. thinking the evidence suggests your sex/race is superior). Alternatively, we may simply lack access to the relevant information to deliberate ethically.

- 2) Inconsistency: Failing to make our moral judgements consistent with each other, as well as failing to align them with our values and changing them appropriately when confronted with new evidence or information.
- 3) Akrasia: Lacking moral motivation to act on moral judgments.
- 4) Innate biases: Evolutionary dispositions towards, for instance, in-group preferences and a fear of out-groups.

I concede there may be many more examples behind such errors, and some of those listed may be understood as a subset of another (for instance, our innate biases may ultimately be thought of as reflecting inconsistencies in our thinking). However, my point here is that it seems to me some combination of these is at the heart of most errors in moral judgements, and therefore the acts that are based on those judgements. Being misinformed or more generally uninformed about reality is probably the biggest culprit: from the dogmatic assertions that there is evidence for race superiority, or for the existence of witches, or for the falsity of others' religion, or for female genital mutilation improving fertility, or that human sacrifice can appease gods – it is difficult to imagine how much violence, discrimination, suffering and uncooperativeness erroneous beliefs or a lack of access to correct information have caused.

Failing to align one's moral judgments with one's values and changing them appropriately when confronted with new evidence or information is a more complex source for errors in moral judgements. It may express itself in myriad ways. For instance: valuing tolerance and freedom, but not realizing how discriminating against a certain minority is inconsistent with those values. Another example may be an inability to perceive that a racial minority behaves a certain way because of complex socio-economic forces rather than anything to do with biology (though this may obviously tie in with having a false belief about a factual matter). The third reason is more straightforward: many believe they should be doing more to help the disadvantaged, for instance, but that belief often fails to translate into action. The final reason is more speculative: as I will show, there is some evidence that we have evolved innate biases we are not aware of, but that drive us to discriminate against individuals we do not consider members of our group. That is, while we might not be explicitly racist, implicit biases may negatively impact how we treats others, leading to intolerance or violence.

It seems to me in all these cases, enhanced cognition would work to reduce the strength of these causes.

- 1) Enhanced cognition would reduce gullibility and augment one's ability to critically evaluate and even uncover factual matters. This would go a long way in reducing the chances of being misinformed or confusing dogma or tradition with evidence.
- 2) Enhanced cognition entails processing information more effectively and consistently, thereby reducing the chances of holding inconsistent values and/or beliefs.
- 3) Enhanced cognition can entail a higher awareness of our self-concepts and personal moral standards, as well as the cognitive dissonance that arises from acting in a way that is inconsistent to them. This heightened awareness, I will argue, would likely increase moral motivation.
- 4) Enhanced cognition of some kind would likely lead to a greater ability to detect or perceive implicit biases within ourselves, leaving us in a better-informed position to stop them from affecting our behaviour immorally.

Framed this way, we can start to see a relationship form between cognition and moral dispositions. In what follows, I suggest there is a convergence of evidence supporting these four claims. Note that, though I discuss empirical evidence for this relationship, the implication remains hypothetical. Namely, my premises are that, if the evidence I present is convincing, and if we agree that augmenting certain cognitive abilities associated with moral dispositions is prudentially valuable for a person⁶, then such an intervention may be a welfarist moral enhancement. Whether such an augmentation would be possible using emerging neurotechnologies is not a possibility I discuss, although the evidence I present suggests "traditional" cognitive enhancement can improve people morally.

3 Evidence for the relationship between cognition and moral dispositions

3.1 A moral Flynn effect

Evidence for the relationship between cognitive ability and moral dispositions can be examined at a macro or individual level. A compelling, albeit controversial, macro-trend worth investigating as a starting point is Steven Pinker's (2011) so-called "Moral Flynn Effect". Pinker (2011) presents an impressive body of evidence showing the world at present is a far more moral

⁶ I presume that such improvements in cognitive ability would be enhancing in the sense that it would tend to increase our chances of leading good lives. Of course, much more can be said about this, but for the time being I take it as uncontroversial that being less gullible, more consistent, and more self-aware tends to improve our lives prudentially-speaking.

place than it has ever been. What is meant here by a more moral world? Broadly, one that is more cooperative, with individuals having larger moral circles and who are less sympathetic to violence. Superficially at least, the truth of this claim is made evident through this passage in *A History of Violence* (Pinker, 2007):

Cruelty as entertainment, human sacrifice to indulge superstition, slavery as a labor-saving device, conquest as the mission statement of government, genocide as a means of acquiring real estate, torture and mutilation as routine punishment, the death penalty for misdemeanors and differences of opinion, assassination as the mechanism of political succession, rape as the spoils of war, pogroms as outlets for frustration, homicide as the major form of conflict resolution—all were unexceptionable features of life for most of human history. But, today, they are rare to nonexistent in the West, far less common elsewhere than they used to be, concealed when they do occur, and widely condemned when they are brought to light.

In fact, despite the atrocities of the 20th century, anyone born in it arguably had a lower chance of meeting a violent death at the hands of another human being than people born in any previous century. Here, this is a broadly statistical claim, and we ought to be wary of the reliability of such figures. But, as a mere illustration, estimates of homicide rates in Oxford were 110 per 100,000 people per year in the 14th century but were less than 1 homicide per 100,000 in mid-20th century London (Pinker 2011). Pinker goes on to present evidence for a similar global trend.

One of the drivers behind this pattern Pinker identifies is reason – specifically, an enhanced capacity for abstract reasoning. Here, Pinker appeals to the Flynn Effect: the finding that people have been getting consistently smarter for at least the past century, especially in regards to their ability to use abstract reasoning (Flynn, 2012). Abstraction from the concrete particulars of our immediate experience is precisely the skill that must be exercised to take the perspectives of others and expand the circle of moral consideration (Pinker 2011). This skill allows one to enter a hypothetical world and explore its consequences and hence is likely to be one of the key ways people rethink their moral commitments. If this skill has played a role in making the world more moral, then the Flynn effect may have caused a corresponding moral Flynn effect.

Needless to say, the relationship between our cognitive ability and moral dispositions is likely more complex than that. Let us explore a number of other avenues that can help shed light on this relationship.

3.2 Intelligence and cooperation

Might being more intelligent dispose individuals toward cooperation? One way to explore this possibility is to look at studies of Prisoners' Dilemma, which highlights the conflict between individual and group rationality (Kuhn, 2014). In these situations, two rational individuals may choose not to cooperate, even if it is in their best interest to do so. In fact, the seemingly rational response in Prisoner's Dilemma is to defect (that is, to confess). Prisoner A cannot trust Prisoner B to cooperate, because Prisoner B has no grounds for trusting Prisoner A to do likewise. And for Prisoner A to cooperate while B defects would result in the worst outcome for A. The frustration of this dilemma, expounded by (Hofstadter, 1985), is that if both prisoners looked down on their dilemma from the "universe's perspective" rather than their own parochial standpoints, it would be clear the best outcome is for both to cooperate. Hofstadter refers to a kind of 'super-rationality' in which both prisoners are certain of the other's rationality.

Pinker (2011) asks whether enhanced intelligence can nudge people towards super-rationality. There is some evidence for this. Burks and colleagues (2009) gave a thousand trainee truck drivers a Matrices IQ test and a sequential Prisoner's Dilemma game. They found participants with better cognitive skills, aside from being more patient in the short and long-term, were more likely to cooperate, even after controlling for age, race, gender, schooling, and income. Smarter truckers were more likely to cooperate on the first move, as well as more likely to respond to cooperation with cooperation.

Another similar piece of evidence comes from Jones (2008) in a study titled "Are Smarter people more cooperative?" This meta-analysis looked at prisoner's dilemma experiments across US universities from 1959 to 2003 and contrasted results with the average SAT scores of those institutions (SATs correlate highly with IQ [Frey & Detterman, 2004]). Jones found the higher an academic institution's mean SAT score, the more its students cooperated in the prisoners' dilemmas. Both of these studies suggest intelligence enhances mutual cooperation, making it more plausible that as a society gets smarter, it is more likely to become more cooperative.

It would be hasty to extrapolate too much from these studies, but they do provide a *prima facie* reason to believe being intelligent would make you more likely to cooperate – at least if you believe your cooperative spirit will be met with a similar one. A society of individuals that can perceive the benefits of cooperation is one that is less likely, all things being equal, to be violent and discriminatory – things that are, by their very nature, obstacles to cooperation.

3.3 Intelligence and being more understanding

Are better reasoning individuals more likely to adopt more understanding policies? Sargent (2004) carried out a study called “Less Thought, More Punishment” and found people with a high need for cognition are more likely to support humane and understanding responses to wrongdoers, even after taking into account factors like age, sex, race, education, income and political orientation. Individuals with a high need for cognition are those who enjoy and engage in effortful cognitive activity, and are marked by a greater susceptibility to changing their attitudes when met with a strong argument, compared to low need for cognition individuals, who typically are not as sensitive to argument strength (Cacioppo, Petty, Feinstein, & Jarvis, 1996).

The Sargent study found a high need for cognition was specifically associated with being less supportive of punitive measures against criminals because individuals with that trait perceive more complex causes for human behaviour – that is, they do not see others in simplistic ways (e.g. good or evil), but as individuals driven by a host of factors, some of them beyond their control. Previous studies support this finding by showing that attributionally complex individuals are more likely to infer “abstract, contemporary, external causes” for behaviour (Fletcher, Danilovics, Fernandez, Peterson, & Reeder, 1986). Attributing criminal behaviour to these kinds of causes would reduce support for stiff penalties, leaving such individuals more disposed to adopting or voting for the kind of policies that are, broadly speaking, kinder and gentler towards, in this context at least, wrongdoers.

Again, it would be rash to extrapolate too much from this study, but it does seem we have a *prima facie* reason to believe being more perceptive of the causes behind others’ actions (an ultimately cognitive ability) would make us generally more understanding and therefore less retributive. The more we are able to perceive and understand the myriad forces that cause people’s actions (their genes, family, early education etc. – all things no one chooses), the more difficult it becomes to hate others for their actions, and the easier it is to be more understanding. In that sense, it seems we have good reason to believe an enhanced perception of this kind would lead to a more moral society.

3.4 Intelligence and being more liberal

There is evidence intelligence correlates positively with classical liberalism. For instance, Kanazawa (2010) analysed two large American datasets and found that in both, intelligence correlates with participants' political liberalism, holding age, sex, race, education, earning, and religion constant. However, what does being classically liberal have to do with being moral?

In general terms, classical liberalism is associated with valuing autonomy and fairness over authority and tradition. People who identify the word "liberal" as representing their politics are more likely to highlight autonomy and fairness over community, authority and purity as values (Haidt & Graham, 2007; Graham, Haidt, & Nosek, 2009). Autonomy and fairness are precisely the kind of values that are likely to lead individuals and, more broadly, societies, to adopt the sorts of norms and policies that reduce discrimination and violence. For instance, a society that respects autonomy over authority is less likely to punish people for disagreeing with the beliefs or ideals of the majority or the authorities. Likewise, a society that highlights fairness is likely to be more concerned with the wellbeing of people who've had a disadvantaged start to life. These are the kinds of values that create the gentler, kinder societies that have emerged from the more violent, less cooperative history that saw authority and tribalism highlighted as values.

Nevertheless, that intelligent people are more likely to be liberal currently looks like a mere correlation. It might be individuals are more likely to cultivate their intelligence if they grow up in the more self-reliant, individualist, autonomy-centric societies that are associated with liberalism. But the evidence suggests the causal arrow starts with intelligence. A study titled "Bright Children Become Enlightened Adults" (Deary, Batty, & Gale, 2008) examined the association between general intelligence (including tests of abstract reasoning) at age 10 and liberal and antitraditional social attitudes at age 30. The study uncovered a latent trait underlying attitudes that are antiracist, proworking women and socially liberal. It found children's IQ at age 10 predicted their endorsement of these attitudes later in life, even holding constant their education, social class and parent's social class. These controls, along with the 20-year long lag, suggest a causal role for intelligence leading to endorsing values associated with classical liberalism.

With these findings in mind, it is worth highlighting that a liberal value system – the kind that highlights individual flourishing as an ultimate good or as society's goal – is arguably the only system justifiable by reason rather than by appeals to authority or tradition. As Pinker (2011) put it: Such a value system "is a product of reason because it can be *justified*: it can be mutually

agreed upon by any community of thinkers who value their own interests and are engaged in reasoned negotiation, whereas communal and authoritarian values are parochial to a tribe or hierarchy.” Indeed, much of the evidence Pinker presents is to illustrate that as diverse people interact in a context of free speech and where history’s failed experiments are critically evaluated, the value systems that evolve are in the direction of liberal humanism. The exceptions that prove the rule are to be found in insular societies lacking a free press and under governmental or clerical repression. It is precisely those societies that cling to tribal, authoritarian and/or religious ideologies.

For a similar conclusion, it is interesting to note that if we were to hypothetically place an individual behind John Rawls’s veil of ignorance – whereby they couldn’t foresee their status, gender, beliefs and so on in a prospective society, and had no set conception of the good – it appears the rational society for them to elect would be liberal in the classical sense; hence Rawls’s (2005) liberal principle of legitimacy. Such a principle would not favour any particular group or creed and would respect individuals’ autonomy and tolerate whatever morally-irrelevant quirks of biology they, for all they know, might end up with. In other words, they would elect a kinder, more understanding, pluralistic and egalitarian society because reason dictates it would be in their best interests to do so. In other words, a prudential concern for oneself demands they elect a moral (i.e. tolerant, understanding, cooperative and therefore less violent) society. Interestingly, the entire thought experiment that is the Rawls’s Original Position is one that fundamentally relies on being able to exercise the capacity of abstract reasoning – that capacity which the Flynn Effect shows has been increasingly enhanced through (at least) recent history. In that sense, enhanced cognition may make more people favour Rawls’s liberal principle.

3.5 Shifts in morality: cultural or cognitive?

An objection worth considering at this stage is whether the above evidence and underlying argument are misattributing changes in our sense of right and wrong to an enhanced capacity among individuals to think abstractly, instead of larger cultural and societal shifts. That is, has racism, for instance, become less acceptable today because more and more people have personally made the cognitive leap necessary to perceive the inconsistency behind ascribing different moral standards based on morally irrelevant racial differences, or is it simply because

people today are born into a “moral world” where most of us have unthinkingly and superficially internalized the idea that racism is wrong.

A prominent theory in moral psychology – moral foundations theory – supports this latter take on morality. Haidt and Graham (2007) argue there are five foundations to morality: harm, fairness, in-group loyalty, authority, and purity. The theory suggests while political liberals have moral intuitions that are primarily based on the first two foundations, conservatives tend to rely on all five foundations equally.

These first two foundations – harm and fairness – relate to protecting the rights and freedoms of individuals and are often referred to as individualizing foundations. The latter three are called binding foundations. They, according to the theory, evolved to preserve social institutions by celebrating and promoting those who sacrifice themselves for the group (in-group loyalty), respect and obey traditions (authority), and rise above their base urges (purity) (Napier & Luguri, 2013).

This understanding of our moral intuitions posits that whether we highlight the first two foundations or all five of them in our moral thinking is determined primarily by environmental forces like cultural differences, rather than cognitive ones. Hence, while liberals highlight harm and fairness in their moral thinking, they do so *not* because of any cognitive analysis on their part, but simply because they live in a “moral world” where the individual is the locus of moral value (Graham et al., 2011). On the other hand, the conservative – and by that is meant the social conservative of the religious right variety – lives in a world that prioritizes the community not the individual, binding the former through perceived proper relationships between, for instance, parent and child, men and women, people and a god.

In that sense, it is precisely through a reliance on these binding foundations that acts of harm and injustice can be directed towards certain individuals or groups, if those individuals or groups threaten the in-group’s notions of authority, loyalty and/or purity. That is, it is people who emphasize these binding principles in their moral intuitions that might set out to exclude those that reject their principles from their moral radius. On the other hand, it is difficult to justify harming or being unjust to others if the two individualizing foundations of aversion to harm and unfairness guide our moral intuitions.

How is this relevant to my argument? Does the moral foundations theory suggest it is false to highlight enhanced abstract reasoning skills as a key driver of moral progress (moral progress

being understood as societies becoming less sympathetic to violence and discrimination and more understanding of differences)?

No. More recent evidence suggests that certain elements of the moral foundations theory are questionable: it is not merely cultural differences that determine whether one relies on the first two foundations or all five, but cognitive ability (Pennycook, Cheyne, Barr, Koehler, & Fugelsang, 2014). For instance, there is evidence the so-called binding foundations are not actually core values in the same way the individualizing ones are. This is marked by evidence that a heightened ability to engage in abstract reasoning tends to diminish the appeal of binding foundations, leaving individuals to highlight the individualizing ones and suggesting they are more foundational values (Napier & Luguri, 2013).

Similarly, Pennycook et al. (2014) provide evidence suggesting differences in cognitive abilities independently predict whether people accept binding moral foundations or not. Specifically, individuals with higher reflective abilities were less likely to view binding foundations as important to their moral thinking, focusing instead on only whether an act was harmful or unfair. They demonstrated this by giving participants two vignettes describing a disgusting act that is generally viewed as morally wrong and asked them to rate how morally wrong each of the scenarios were. The vignettes – based on scenarios of incest and zoophilia – were designed to specifically exclude any care or fairness-based violations of moral values, and to elicit particularly strong disgust-based intuitive responses, as well as emotionally driven responses that are resistant to reasoned persuasion.

The authors suggest the reason for their findings is that participants who are less able to represent abstract concepts mentally (i.e. had lower cognitive abilities in that respect) were more likely to confuse feelings (e.g. disgust) arising from being prompted about binding foundations as indicating moral significance.

This hypothesis is in line with the finding that individuals in abstract mind-sets tend to prioritize individualizing foundations while devaluing bindings ones in their moral judgments (Napier & Luguri, 2013). By priming participants experimentally into either an abstract or concrete mindset and then measuring how much weight they placed on the five moral foundations, Napier and Luguri demonstrated when individuals were thinking abstractly, both liberals and conservatives increased their valuations of individualizing foundations, and decreased their valuations of the binding foundations, as compared to when they were thinking concretely (that is, conservatives came across as liberals when thinking abstractly).

How is inducing an abstract mindset relevant? Being in an abstract mindset refers to focusing on core features consistent across situations. This is in contrast to a concrete mindset, where one is less likely to make a distinction between core and secondary factors. Hence an abstract mindset is associated with making higher level core values more salient. Abstract thinking allows individuals to consider the particularities of any one experience and extract evaluation-relevant information that is stable across time, contexts, and relationships (Ledgerwood & Trope, 2011). The fact that binding foundations are devalued when in this mindset (among both conservatives and liberals) suggests they are not as foundational as the individualizing ones.

These studies convey three things: a concern with harm and fairness represent higher level core values; cognitive ability, rather than environmental factors like culture, plays a bigger role in driving our moral values; more reflective individuals with a higher capacity for abstract thinking are likely to only highlight these individualizing foundations (which is precisely what happens when we think abstractly from behind the veil of ignorance: it seems irrational to choose a basic structure based on in-group loyalty, authority or purity).

In other words, more reflective people, who can think in abstract ways, highlight only values that make it difficult to justify harming or excluding others in the name of loyalty, authority or parochial ideas of purity. Such individuals are more likely to focus on what is actually relevant to our most fundamental moral intuitions precisely because they can abstract across the particularities of given situation or experience. All this suggests that enhanced abstract reasoning leads to better moral judgments – or at least leads to valuing the sorts of principles (namely, autonomy and fairness) that are likely to correspond with what most reasonable people would agree is a more moral society.

Altogether, the evidence presented here makes a case for believing improved cognitive abilities could act as moral enhancements – at least in terms of having better moral judgements, and only so long as such improved abilities are generally prudentially beneficial for those who undergo them. However, improved moral judgements are arguably only one part of what moral enhancement entails.

4. Moral judgments versus moral motivation

Are moral judgements sufficient to motivate us to behave morally? This is a contested topic. For instance, DeGrazia (2014) argues moral judgements comprise only half the story. Moral

behaviour – which is presumably the end goal of moral enhancement – is not only the product of having moral judgments, but also moral motivation: that is, a drive to do what is held to be right. In that sense, as he puts it: moral insight + moral motivation → moral behaviour.

This has intuitive appeal. Most of us are familiar with holding all sorts of moral beliefs while failing time and again to act on them. We seem to care in principle, but not in practice. But is that incongruity because, despite our proclamations, we don't truly hold these moral beliefs? Is that what causes the lack of motivation? Or is it that moral motivation is separate from our beliefs, perhaps arising from other desires?

Philosophers and moral psychologists continue to disagree about the relationship between having a moral judgment and having moral motivation. Internalism is the doctrine that a moral judgment necessarily comes with a motivation to act on it – one that is inherent to holding that very judgment (Rosati, 2016). In contrast, externalism holds moral motivation as something separate that arises from other desires that are not necessarily tied to the moral judgment.

On the one hand, there is evidence from individuals with injuries to the ventromedial cortex that suggests internalism is false: these individuals appear to suffer from “acquired sociopathy”, in that they appear to retain a sense of right and wrong and hold moral judgments, but lack any inclination to actually act on them (Roskies, 2003). Roskies sees such individuals as walking counter-examples to the claim that moral judgments necessarily entail moral motivations to act on them.

On the other hand, Cholbi (2006) argues it is not clear such individuals actually hold the moral judgments they profess to hold. Moreover, as Björklund and colleagues (2012) demonstrate, there are a multitude of ways of framing internalism, from simple internalism – i.e. if a person judges she morally ought to do something, then she is necessarily at least somewhat motivated to do it – to conditional internalism, which adds the requisite that the moral agent in question is psychologically normal and/or morally perceptive.

For the purposes of my argument, it suffices to show that whether some form of internalism is true or not, what matters is to demonstrate moral judgments and motivations (regardless of the relationship they share) are subject to brain states that are at least partly cognitive in nature – and hence enhanceable in that way. I believe I have already showed this for moral judgments. If internalism is true, then enhancing the cognitive abilities that lead to moral insight should be sufficient to lead to moral behaviour automatically. If externalism is true, we need to consider how moral motivation might be enhanced. I argue the best way to approach moral motivation

enhancement is also through cognitive enhancement of some kind. Before elaborating on this, note this suggestion Crockett (2014) makes:

DeGrazia distinguishes between the enhancement of moral motivation, moral cognition and moral behaviour. From a neuroscientific perspective, the evidence so far suggests that targeting moral motivation may be the most promising avenue for promoting moral behaviour.

In support of this, Crockett notes evidence that mentally healthy individuals given the anti-depressant citalopram appear to have their moral behaviour affected through a motivational channel, by increasing harm aversion (Crockett, Clark, Hauser, & Robbins, 2010).

As I've argued, such an approach may raise concerns about infringing on autonomy and individuality (Harris, 2014; Schaefer 2015). This approach of directly modifying emotions is also potentially subject to criticisms of moral perfectionism: if moral enhancement is to be executed at a societal or global level, it entails states or authorities would have to take sides in the debate on what constitutes a good character (Sparrow, 2014). The problem here is the elitism entailed in moral perfectionism: in a pluralistic society, who is to say what a good character is? On the one hand, it seems uncontroversial that a good character is not a murderous one, but exactly how altruistic and sympathetic is a good character? How are we to determine what constitutes a finely tuned moral character (presuming that was possible)? These questions are not unresolvable and I do not raise them to suggest direct moral enhancement is necessarily impermissible. Instead, my goal has been to offer an alternative approach to sidestep these concerns.

4.1 Moral motivation enhancement

One way to resolve these concerns about directly modifying motivation is to approach it from a different, more cognitive angle. A primary way through which people exert control over themselves is through an internal reward system that acts on an individual's self-concept – that is, the way individuals perceive themselves (Aronson, 1969; Baumeister, 1998). It has been shown that people have strong beliefs in their own morality, and that they strive to maintain that aspect of their self-concept (Greenwald, 1980; Griffin & Ross, 1991; Sanitioso, Kunda, & Fong, 1990). This suggests when an individual fails to live up to their own internal standards, they feel pressured to update their self-concept negatively, which is highly aversive. The tension created from wanting to avoid negatively updating one's self-concept is in turn highly

motivating. People will be driven to maintain a positive self-concept even if it involves sacrificing financial gains and/or investing effort (Aronson & Carlsmith, 1962; Harris, Mussen, & Rutherford, 1976).

This opens up the possibility of approaching moral motivation enhancement in this indirect way: by augmenting one's attention to their moral standards and how they tie in to their self-concept. In a sense, this can be understood as enhancing one's integrity: boosting self-awareness (an attentional capacity) so one is better able to perceive whether their actions are in line with their beliefs. A heightened awareness of one's self-concept (e.g. "I am fair") and the moral standards it entails (e.g. "I don't cheat"), along with a greater ability to perceive whether an action is in line with those standards or not (all ultimately cognitive skills), would seem to lead to enhanced moral motivation.

One study lending evidence to this proposal looked at the relationship between acting honestly and attending to one's moral standards, as well as the ease with which a situation made it easy to justify dishonesty (Mazar, Amir, & Ariely, 2008). Through a series of six experiments that variously accentuated the attention of participants to their standards of honesty, the researchers found dishonesty increased as attention to personal standards of honesty decreased. Their results suggest making people more mindful by increasing their attention to their honesty standards curbs dishonesty. That is, greater self-awareness appears to motivate individuals to act in a way consistent with their moral beliefs.

Note also this approach to moral motivation enhancement can also reduce harmful acts arising from innate and implicit biases that dispose individuals to in-group favouritism or out-group aggression (the fourth explanation I posited for errors in moral judgment). A classic example of this is implicit racial bias. Implicit pro-white/anti-black bias appears to be evident in white American children from as early as six years old, remaining stable among 10 year olds and adults, with only explicit race biases declining with age (Baron & Banaji, 2006). If cognitive enhancement is harnessed so it augments self-awareness or the ability to attend to or perceive such biases, individuals could be in a better position to take note of them and, if not change them, at least limit how they impact their behaviours.

On the other hand, while this approach to moral motivation enhancement might side step the problems associated with direct emotion manipulation and the moral perfectionism it may entail, it also presupposes that whoever is undergoing the intervention already has sound moral judgments. It is conceivable that for some people, the only reason why they are not waging holy wars or martyring themselves for some occult cause is simply because they lack motivation. They

lack the motivation to act on their moral judgments – and a good thing too! However, it seems to me the potential unwanted effects of purely moral motivation enhancement can be avoided by approaching moral enhancement through boosting certain cognitive capacities. To the extent that some forms of cognitive enhancement can augment reasoning abilities as well as self-awareness/attention to our self-concepts and how our actions impact them, then it seems we have hit two birds with one stone: simultaneously enhancing moral reasoning (and thereby, presumably, our ability to reach sound moral judgments), as well as moral motivation (presuming that motivation is divorced, or can be, from moral judgments).

Such would be an approach to moral enhancement that can be supported by, both, those wary of directly modulating moral motivation (Harris, 2014; Schaefer 2015; Sparrow, 2014) and those seemingly unfazed by it (Crockett 2014; DeGrazia 2014; Douglas 2013; Persson & Savulescu, 2008, 2012).

Of course, currently it is not clear what it would take for biomedical cognitive enhancers to work to that end. As we will see in the next chapter, there is no evidence the pharmaceutical interventions presently under discussion boost abstract reasoning or self-awareness of the sort described here. As such, this has not been an argument for using, say, Modafinil, to boost moral behaviour. There is no evidence such a drug would have any impact on moral behaviour. Instead, it has been an argument for how it might be best to approach the prospect of direct moral enhancement in welfarist terms.

5. Conclusion

I initially argued that we may wish to distinguish welfarist from moral enhancement; however, using a number of thought experiments I highlighted how a welfarist evaluation of moral technologies adds a crucial dimension to understanding different types of moral enhancers (ones that are advantageous to users and ones that aren't). In that way, the welfarist account provides a useful framework for thinking about the implications of different moral technologies or interventions by highlighting their impact on the well-being of the users of such technologies. I then argued such an approach may favour an indirect form of moral improvement through cognitive enhancement. I posited that errors in moral judgements can be understood as reflecting cognitive deficits of some kind or other: by enhancing cognition, it seems these sources of errors in moral judgements would be greatly diminished. Here I have presented

converging evidence suggesting that improving certain aspects of cognition would likely lead to reducing the chances of holding the sorts of moral judgments that most reasonable people would agree lead to immoral behaviour.

On other hand, what about people who have agreeable moral judgements, but lack moral motivation? Again, regardless of our account of the nature of moral motivation, it seems like whether one lacks it or not depends on cognitive states that can in principle be enhanced. Given the motivating tension arising from acting inconsistently with our self-concepts, augmenting our ability to attend to those self-concepts and the moral standards they entail, would likely boost moral motivation. This heightened self-awareness would also curb the power implicit biases may hold over us, simply by exposing them to our conscious minds.

It is also interesting to note here that, based on the conclusions reached in the last chapter on the extended mind, it may be that various changes to certain aspects of our environment may be moral enhancements in the welfarist sense. If such altered external cues come to form parts of the psychology of an individual, and impact them in a way that improves their moral deliberations or motivates them to act morally, they may be instances of welfarist moral enhancement (so long as they tend to contribute to one's well-being).

At this stage, I consider my examination of the welfarist account sufficient to begin deploying it, having highlighted its numerous implications and limitations by engaging with the account's definition of enhancement and the arguments of its critics, by exploring the implications of EMT for the account, and by investigating the seeming conflict it holds with moral enhancement. In the next chapter, I move on to consider how useful this account will be for practical ethicists and policy-makers seeking to deploy it for regulatory purposes. I do this by exploring what the welfarist account can tell us about putative "smart drugs" currently in use. Do they, in fact, improve people's well-being? As we will see, answering this question requires delving into and challenging a number of presumptions in the associated literature.

Chapter 5

Smart Drugs and Welfarist Enhancement

In chapter one, I argued the debate over enhancements should distinguish between concrete and abstract ethical questions. So far, my thesis has focused mainly on conceptual issues and abstract questions raised by the welfarist account enhancement, exploring the relevant set of circumstances, the boundaries of our psychologies, and how the account relates to moral enhancement. My investigation, therefore, has remained detached from real world questions about how the welfarist account would interact with a concrete case of a putative enhancer. This chapter serves to illustrate this interaction; specifically, it serves to illustrate how practical ethicists and policy makers can use the account to begin thinking about the ethics of a transmedical intervention. To that extent, my goal is not to provide conclusions about the permissibility of a given intervention, but to give an example of deploying the account.

I do this here by examining the case of so-called smart drugs: a paradigmatic case that has been massively discussed academically and popularly, normally in the context of these drugs being a functionalist, transmedical intervention. As I mentioned in the introduction to this thesis, I focus on this topic specifically (as opposed to moral or some other type of enhancement) because, as I will show, it is a prime example of how unstated presumptions, poor conceptual clarity, inadequate reliance on empirical data, and a propensity for hype get in the way of clear thinking about enhancement. That is, it is a prime example of ignoring the methodological guidelines for this thesis, and more importantly, of poorly engaging in practical ethics.

In a manner, this chapter aims to give an example of how to combat this trend: before getting to the question of whether smart drugs are welfarist enhancers, I first state what I think the key presumptions in the associated literature are (section one), clarifying how the literature tends to define cognitive enhancement (section two), before then investigating those presumptions and arguing for why they are unfounded (section three) – a task that is largely carried out through a rigorous investigation of empirical evidence. I then lay out the costs that come with these presumptions (section four), note some positive, more recent shifts in the literature (section five), and highlight remaining conceptual issues that the literature has not dealt with (section six). It is only at this stage that I begin exploring what the welfarist account would make of smart drugs (section seven), having now properly better understood what these are along

with appreciating the limited and misleading evidence related to them. In doing so, I also show how the account can be used to systematically think about any intervention to the body or mind.

Note that aside from section seven of this chapter, I will be referring to these drugs here in the functionalist, transmedical sense. The reason for this temporary shift is because I will be engaging with literature that characterizes smart-drugs as putative enhancers in the transmedical sense.

1. Recurring presumptions about smart drugs¹

The debate on biomedical cognitive enhancement, in the functionalist sense, has been massive. In fact, the sheer volume of discourse about this prospective technology has led some to wonder whether the ethical investigation is too far ahead of the technology (e.g. Quednow, 2010). What is fuelling this frenzy of deliberation about such a specific prospective technology? In the context of smart drugs, or pharmacological cognitive enhancements (PCE), which have received the bulk of the attention in the associated literature, I argue that a number of key presumptions act to propel this flurry of debate. These can be summarized as follows:

- That evidence exists showing certain drugs boost cognition among the cognitively normal.
- That it makes neurological sense to talk about pharmacologically enhancing cognitive abilities separately from non-cognitive facets of our psychology, like mood and motivation.
- That an ethical analysis of cognitive enhancement specifically is therefore pertinent.
- All the more so due to the evidence of widespread and increasing PCE use, especially among students.

As I will illustrate below, some combination of these four presumptions is at the heart of this debate. My goal here is to expose these as unwarranted. I will first argue that, in fact, there is no consistent evidence that the drugs commonly discussed – such as methylphenidate (Ritalin), mixed amphetamine salts (Adderall) and modafinil (Provigil) – actually augment specifically cognitive capacities among the healthy and non-sleep deprived. I will then provide evidence that

¹ A modified version of my writing from here until section seven of this chapter has been published while working on my thesis: Zohny, H. (2015). The myth of cognitive enhancement drugs. *Neuroethics*, 8(3), 257-269.

these drugs appear to be acting on non-cognitive brain states, and that attempts to separate the cognitive from the non-cognitive (such as mood and motivation) are artificial and neurologically misinformed.

Furthermore, I will argue that maintaining any such artificial divide between PCE and pharmacological non-cognitive enhancement (PNCE) in the name of highlighting the core ethical issues raised by the former, such as using PCE in academic examinations, is unfounded: PNCE that boost subjective states such as confidence and motivation are equally relevant in such contexts. In fact, it is not clear how PCE *per se* raises any ethical questions that are distinct from some broader enhancement of brain states. I will then argue that a recurring justification for this debate – that the use of PCE is widespread and likely to increase – lacks evidence and is based on scant and unreliable data. Finally, I reflect on some of the costs and implications of these presumptions.

The goal here is not to conclude that the debate itself is unfounded. Instead, the aim is to highlight the importance of a less hype-driven conversation that is instead rooted in greater conceptual clarity and in the actual empirical evidence for PCE efficacy and prevalence. Doing this first sets the groundwork for then examining what the welfarist account would make of these drugs.

2. Cognitive enhancement in the literature

Before proceeding, it is important to be clear about what exactly is meant by cognitive enhancement in the context of this debate. While I have argued in chapter three that cognitive enhancement may involve changes to our environment, this is not typically considered in the literature, and the term is often used interchangeably with other “internalist” descriptions such as neuroenhancement – often used in an intentionally interchangeable manner with cognitive enhancement (e.g. Lucke, Bell, Partridge, & Hall, 2011) – as well as with terms like “brain doping” and “cosmetic neurology” (e.g. Cakic, 2009).

Nevertheless, a clear domain for cognitive enhancement is apparent in the literature, and that domain is rooted in that aspect of our psychology associated with the organization and processing of information. It is such traits as perception, attention, understanding, memory, and reasoning that are associated with cognition (Bostrom & Sandberg, 2009). Hence in a volume titled *Cognitive Enhancement: An Interdisciplinary Perspective* (Hildt & Franke, 2013), it is

characterized as increasing “information-processing functions such as learning, planning, concept formation, perception, attention, memory, reasoning and problem solving” (Hiltdt, 2013). Despite the seeming interchangeability of cognitive enhancement with neuroenhancement, Kipke (2013) argues that the latter term can encompass the enhancement of both cognitive and non-cognitive properties (such as mood and motivation), while cognitive enhancement refers only to specific enhancements that aim to improve cognitive function. PCE can be seen as one form of cognitive enhancement, and is commonly considered in terms of improving cognitive capacities like attention and memory through the use of pharmacological agents like methylphenidate, Adderall and modafinil (Singh, Bard, & Jackson, 2014; Pustovrh & Mali, 2013).

While not necessarily articulated in these characterizations of cognitive enhancement, such interventions are designated as actual enhancements only to the extent that they improve capacities in some way other than by repairing or remedying a specific dysfunction. Thus Bostrom and Sandberg elaborate on cognitive enhancement as an intervention that improves the performance of some cognitive subsystem without correcting some specific, identifiable pathology or dysfunction of that subsystem (Bostrom & Sandberg, 2009). In that sense, the target of these interventions are cognitively normal individuals, at least to the extent that we can agree on what “normal” currently refers to.

This emphasis on the information processing character of cognition is often put into context by contrasting it with non-cognitive enhancements that impact more subjective brain states like mood and motivation, which are deemed to be a different of kind intervention. Hence, a volume like *Enhancing Human Capacities* (Savulescu, ter Meulen, Kahane, 2011) dedicates separate sections for cognitive enhancement and mood enhancement. While I will argue that attempts at making a clear-cut separation of our information processing faculties from the non-cognitive facets of our psychology are misguided, we can at least move on now with an idea of how cognitive enhancement is generally characterized in the literature. For the purpose of PCE, it is an intervention that aims to pharmacologically boost already healthily functioning information processing faculties, as opposed to more subjective facets of our psychologies, such as our emotions and dispositions.

3. Four presumptions

3.1 Presumption #1: Cognitive enhancement drugs already exist

As stated above, much of the debate in the literature on PCE presumes that these drugs exist, and that their effects on our psychology are specifically cognitive in nature. Take this declaration by Buchanan: “Drugs designed to treat the symptoms of Alzheimer’s dementia (including Aricept), to treat attention deficit disorder with hyperactivity (including Ritalin and Adderall), and to treat narcolepsy (Provigil) have all been shown to improve cognitive function in the cognitively normal” (Buchanan, 2011b).

Similarly, Savulescu asserts: “Already, drugs are being used to improve cognitive performance in the normal range. Modafinil is a new class of drug originally developed for narcolepsy... It improves executive function, wakefulness, and working memory” (Savulescu, 2011). Speaking of Adderall and Ritalin, Greely and colleagues (2008) note that “these drugs increase executive functions in patients and most healthy normal people, improving their abilities to focus their attention, manipulate information in working memory and flexibly control their responses.”

These are deemed uncontroversial proclamations, so much so that they often go unreferenced. Here I argue that even in instances where empirical support is cited, the overall evidence is far from conclusive that these drugs actually improve cognitive function in the cognitively normal.²

For instance, a survey of more than fifty experiments on the effects of drugs that treat ADHD on a wide array of cognitive functions – such as memory and executive function – among healthy young adults found a roughly even mixture of significant enhancement effects and null effects (Smith & Farah, 2011). This was particularly prevalent in the case of the stimulants’ impact on working memory. Moreover, all of the null results the review uncovered were part of studies reporting on multiple tasks. The fact that all studies that measured effects on single-task experiments found an enhancing effect suggests that at least some single-task studies with null results have gone unreported (Smith & Farah, 2011).

This should serve as a reminder that the empirical evidence for cognitive enhancement suffers from the same problems facing experimental psychology and clinical trial research in general. Most studies on cognitive enhancement are too small and underpowered to find anything but

² Recall that I am referring to cognitive enhancement here in the functionalist, transmedical sense.

large effects (Farah, Smith, Ilieva, & Hamilton, 2014). It is also impossible to gauge how many null results go unreported since these studies are not normally registered in advance.

So what does the evidence from systematic reviews and meta-analyses actually suggest? One such review suggests that expectations regarding the effectiveness of methylphenidate (Ritalin) and modafinil (Provigil) exceed their actual effects (Repantis et al., 2010). With regard to Ritalin, the reviewers were not able to provide sufficient evidence of positive effects in healthy individuals from objective tests. While the reviewers reported some positive effects from modafinil, they stressed that these remained equivocal and more studies were necessary.

In regards to modafinil specifically, Chamberlain and colleagues (2011) reviewed studies in which the Cambridge Neuropsychological Test Automated Battery (Sharma, 2013) had been used to assess its effects on patients and healthy control participants. They found that its use among healthy participants showed no effects on sustained attention or attentional set shifting. On five measures for which sufficient data from previous studies were available, they found no evidence overall for significant effects of modafinil among healthy participants.

Likewise, Farah et al. note that while there is clear evidence that modafinil improves executive function and memory for sleep-deprived individuals, perhaps making it a kind of performance maintainer, findings are mixed with well-rested adults, including occasional cases of impairing effects (Farah et al., 2014). Notably, two recent studies not covered by these reviews found no evidence that modafinil boosts cognition, but that it actually slows down response time, potentially also hindering creativity (Mohamed & Lewis, 2014; Mohamed, 2014b).³

Another review concluded that while studies in non-ADHD adults suggest that stimulants may improve retention of previously acquired information and possibly facilitate memory consolidation, they do not promote the acquisition of new information, and may actually impair performance on tasks that require adaptation, flexibility and planning (Advokat, 2010).

As for the effects of acetylcholinesterase inhibitors such as donepezil (Aricept) on healthy individuals, mentioned by Buchanan above, a systematic review found that “the few existing studies, mostly about donepezil, provide no consistent evidence for a neuroenhancement effect” (Repantis, Laisney, & Heuser, 2010). In fact, two of the studies surveyed reported transient negative effects on episodic memory (Beglinger et al., 2004, 2005).

³ Further complicating matters, a more recent systematic review on modafinil (Battleday & Brem, 2015) concluded the drug may provide some benefit to cognition, especially executive functions.

Regarding mixed amphetamine salts (Adderall), a study examining its effects on 13 different measures of cognitive performance – including intelligence, scholastic achievement, inhibitory control, episodic memory and working memory – failed to find any evidence of reliable enhancement among healthy participants (Ilieva, Boland, & Farah, 2013). The study is notable for avoiding the insufficient statistical power of similar, previous studies, using a sample size that would allow for 95% power to detect any medium-sized effects on any single measure.

More generally, there is evidence that while these substances can augment cognition in subjects with low baseline capacities, these effects disappear among subjects with higher baseline performance, and may even lead to a deterioration in performance (Mattay et al., 2000). Similar findings were reported by a review from the British Academy of Medical Sciences. While some drugs were reported to improve cognition in impaired individuals, the evidence for any enhancement among healthy individuals was found to be very limited, normally being observed only in laboratory conditions, not everyday life (British Medical Association, 2007).

At this point it is clear that, given the evidence, it does not seem justifiable to state, as some of the authors quoted above so confidently have, that there are cognitive enhancing drugs out there at the moment. Perhaps the evidence suggests, given the more obvious impact these drugs have on sleep-deprived individuals, that they act as “performance maintainers”, as has been recently suggested (Kjærsgaard, 2015). But among healthy, non-sleep deprived individuals, whether it’s Aricept, Ritalin, Adderall or Provigil, the evidence for enhancement from systematic reviews is not promising. Authors that cite one or two studies that report an enhancing effect is insufficient to justify their claims about these drugs.

To that extent, the first presumption partially driving this debate – that there is conclusive evidence these drugs enhance cognition amongst the cognitively normal – appears to be a myth. What this says about their prospect as welfarist enhancements will be crucial: given the evidence for trade-offs and impaired performance in certain cases, they may actually be disabling changes to our psychologies. I explore this in more detail in section seven.

3.2 Presumption #2: Cognitive enhancement is separable from non-cognitive enhancement

This lack of consistent evidence highlighted above is difficult to reconcile with anecdotal reports from healthy individuals who insist that, at the very least, *something* about these drugs seems

to make them study better or work harder. What are we to make of this? In this section I argue that these drugs appear to be acting, if at all, on non-cognitive brain states. This will help illustrate a further point: in the context of enhancement via pharmacological means, it is misleading to characterize the cognitive as separate or separable from the non-cognitive.

A study mentioned above is notable for measuring the subjective effects of Adderall (Ilieva et al., 2013). It found that participants tended to *believe* their performance was enhanced when using the drug, compared to when on a placebo. Similarly for modafinil, while it is not clear that it actually improves objective performance, it appears to induce the subjective impression of better functioning (Brukamp, 2013), with some evidence that it may induce a sense of overconfidence (Baranski, Pigeau, Dinich, & Jacobs, 2004b), all the more so in individuals who have been sleep deprived for 64 hours (Baranski & Pigeau, 1997). In fact, one notable study found that administering modafinil to healthy young volunteers who were not sleep deprived had no effect on cognitive performance, but was associated with significant changes in mood (Randall et al., 2003).

Findings like these have led Farah and colleagues (2014) to suggest that, it may be the non-cognitive effects of these drugs that are actually enhancing performance.

Consider these quotes from university students who have used prescription stimulant drugs for nonmedical purposes (Vrecko, 2013):

“Everything seems better, and more doable.”

“When I’m not on it I’m usually pretty relaxed about things. [. . .] When I’m on it it’s like more of a sense of urgency.”

“It just got to where I felt like if I was staring at something I just couldn’t take my eyes away from it—it made studying more interesting.”

“You start to feel such a connection to what you’re working on. It’s almost like you fall in love with it—there’s nothing else you’d rather be doing!”

Given these sentiments, Scott Vrecko (2013) questions just how “cognitive” cognitive enhancement really is. Based on his interviews, he has identified four recurring non-cognitive themes related to subjects’ experiences with these drugs. He calls them: “feeling up”, “drivenness”, “interestedness”, and “enjoyment.” Each of these refer to affective and motivational states presumably induced by stimulant drugs.

This suggests that even if there is no evidence that these drugs enhance cognitive ability, they may still be inducing certain moods and motivational states in users that enable them to make the most of their standard, unenhanced cognitive capacities, improving their performance in that way.

If so, it seems misleading to continue categorizing these drugs as forms of pharmaceutical *cognitive* enhancement. But even if new evidence were to conclusively show that they do objectively enhance cognitive ability specifically, I want to argue that it is neuroscientifically uninformed, if not also conceptually confused, to discuss the cognitive as separate from the non-cognitive. It is not clear what it would mean to selectively modify cognition pharmacologically but not mood, or vice versa. Superficially at least, it would seem that pharmacological interventions that improve mood may imply boosting motivation, which can surely affect how we utilize our cognitive abilities. Similarly, augmenting cognition to become more alert or mindful may entail feeling more serene or better more generally (Hildt, 2013).

More specifically, neuroscientific research suggests these two facets of our psychology are neurally intertwined, with emotion affecting cognition, and vice versa. For instance, when we are sad, concentration can be more difficult and we become more selective in what we recall (Dolan, 2002). Similarly, the emotional importance of an event strongly impacts whether it is consolidated into episodic memory or not (Hamann, Ely, Grafton, & Kilts, 1999). Likewise, feared stimuli (e.g. the sight of a snake) has a significantly enhancing effect on how rapidly we can detect it, with emotions more generally playing a crucial role in driving attention (Ohman, Flykt, & Esteves, 2001). These examples serve to illustrate the integrated nature of the cognitive and non-cognitive aspects of our psychologies.

Ultimately, the idea that we can pharmacologically alter the cognitive without also impacting the non-cognitive can make sense only through a simplistic understanding of how the brain works, with the limbic system functionally specializing in processing emotion and the cerebral cortex dedicated to cognitive capacities. The reality is more complex than this “one-area/one function” viewpoint (Pessoa, 2008). The picture emerging from connectivity data suggests a far more integrative system (Gray, Braver, & Raichle, 2002). Even regions that were once thought to be purely affective in their specialization, such as the amygdala, hypothalamus and nucleus accumbens, appear to function more as connectivity hubs, integrating information from all over the brain rather than merely processing emotion (Pessoa, 2008). For instance, while the amygdala is normally described in terms of affective functions, it is also closely linked to cognitive capacities such as associative learning and attention (Holland & Gallagher, 1999).

Likewise, brain regions commonly viewed as purely cognitive areas, such as the lateral prefrontal cortex (LPFC), also appear to play more integrative roles rather than being specialized in cognition. For example, functional studies of the LPFC suggest that emotion and cognition are strongly integrated there (Gray et al., 2002; Perlstein, Elbert, & Stenger, 2002). In fact, evidence from monkey studies show the LPFC integrates not just cognition and emotion, but motivation as well (Watanabe, 1990, 1996)

Overall, the emerging picture from neuroscience suggests a dynamic interplay between interconnected neural systems involved in non-cognitive and cognitive functions (Dolcos, Iordan, & Dolcos, 2011). While this more nuanced understanding of brain processes suggests that the notion of purely enhancing cognition pharmacologically is neurologically misinformed, it does help make sense of individuals' subjective experiences with these drugs. The relationship between how we utilize our attention and working memory is bundled up with how, say, interesting and enjoyable we perceive what we are doing to be. Alter this latter subjective component (as these drugs appear to), and it may well indirectly impact how we use and feel about our standard cognitive abilities.

It may be countered that this means these drugs are cognitive enhancers nonetheless: if a drug evokes a mood that ultimately leads to sustained effort at a cognitive task, what does it matter whether that effort is thanks to better concentration or simply a more interested disposition? The result, in terms of performance, is the same. However, based on how cognitive enhancement is typically characterized in the literature (see section two), what seems of concern is improving cognitive function specifically rather than performance. Ultimately, whether it is function or performance that is of concern, functionalist accounts need to clarify this, and the way to do that is, as I've argued here, acknowledge the inter-related functioning of our cognitive and non-cognitive capacities. As we will see, for the welfarist account of enhancement, which is not concerned with functioning in itself but only with changes to our psychologies more generally, there is less scope for such confusion.

Nevertheless, it might be argued that artificially separating cognitive capacities from subjective states in the context of enhancement is worthwhile for other reasons. For instance, Hildt (2013) argues that an ethical analysis separately dedicated to cognitive enhancement and mood enhancement is helpful because of the significant differences in the contexts in which these are sought, such as the aims motivating their use, and the consequences entailed by each. Unlike mood enhancement, it might be argued, PCE is more likely to be sought in the context of academic or professional settings, with an aim of being a better performer in a competitive

setting. This would in turn have important consequences for our schooling systems and, ultimately, the workforce and economy. Hence, an emphasis on the ethical implications of PCE specifically, it can be argued, is warranted. In the next section, I argue that it isn't.

3.3 Presumption #3: There are ethical issues that are uniquely raised by cognitive enhancements

This presumption is often implicit in the literature. The widespread use of the term cognitive enhancement in the debate suggests that the ethical issues raised by these pharmacological interventions are in some way unique compared to the non-cognitive enhancement of mood or motivation (Kipke, 2013).

Hence, Bostrom and Sandberg (2009) specify that cognitive enhancement raises ethical issues by interacting with notions of authenticity, the good life, and the role of medicine. Goodman (2010) questions how cognition-enhancing drugs might be construed (or misconstrued) as cheating or diminishing of our personal accomplishments. Cakic (2009) reviews the ethical and pragmatic implications of cognitive enhancing drugs in academia. In all three cases, and many more (e.g. (Bostrom & Roache, 2010; Hauskeller, 2013; Hildt & Franke, 2013; Tännsjö, 2009), why is the concern focused on cognitive enhancement specifically?

Note first that it is not being suggested that these authors claim that cognitive enhancement necessarily stands as uniquely interesting or different in the ethical questions it raises. But rather, this seems to be the overall implication coming from the associated literature. Suggestive evidence for this is reflected in the 821,000 results Google Scholar uncovers when searching for "cognitive enhancement", compared to the 186,000 under "mood enhancement".

As noted above, Hildt (2013) suggests that it is because of the unique contexts, reasons and consequences posed by cognitive enhancement in particular, as opposed to non-cognitive enhancement, that has resulted in this special attention. And yet, it is not clear that these reasons hold up when we investigate the specific recurring ethical issues of these two types of enhancement.

To demonstrate this, let's examine these ethical issues and see if perhaps cognitive enhancement raises different ones, or raises them in a particular way that demands special attention.

To an extent, such an analysis has been carried out by Kipke (2013), who has identified nine recurring ethical issues, demonstrating how these overlap in similarly relevant ways whether discussing cognitive or non-cognitive enhancement. These are: efficacy and safety; personal authenticity; quality of happiness; development of character and self-awareness; accountability of achievement; fairness in competition; social pressure; changes in values and the idea of humankind.

To illustrate, we can ask whether authenticity is an issue that is relevant only to one domain of psychological enhancement in a unique way. Bostrom and Sandberg (2009) question whether making cognitive abilities for sale in the form of a pill would make them less genuine. Yet this same question certainly seems equally relevant when considering mood enhancement (Berghmans, ter Meulen, Malizia, & Vos, 2011). Is a cheerful mood induced by a pill a less genuine expression of oneself? The goal here is not to answer these questions, but to show that they apply to the prospect of enhancing cognitive and non-cognitive traits equally.

Similarly, Goodman (2010) asks whether the use of cognitive enhancing drugs unnaturally cheapens accomplishments. But it is difficult to see why this might be an issue specific to cognition and not to other aspects of our psychology.⁴ As Kipke (2013) points out, if someone improves their social and communicative skills through a mood enhancer and ends up accomplishing something great (say, peace between two warring nations), the same question of attributability arises. It is not clear that the question of attributing accomplishments is only relevant when cognitive power is used to find solutions, as opposed to non-cognitive skills, like having a friendly disposition or being in an empathic state of mind.

Kipke (2013) also questions whether cognitive enhancers pose a particular problem for fairness in competition (and social justice more generally) in a way that non-cognitive enhancers do not. Are social opportunities primarily determined by cognitive abilities? Authors often focus on pharmacologically augmented IQs as a key example that may lead to unfairness in, say, securing a job. But competition over such opportunities surely extends beyond cognitive capacities. True, a successful professional life requires cognitive skills, but also social, communicative and emotional skills. In fact, individuals who lack self-confidence, self-control, or empathy don't tend to have much success – at least, these more non-cognitive traits appear crucial to success (Goleman, Boyatzis, & McKee, 2013; Tangney, Baumeister, & Boone, 2004). In that sense, if

⁴ Again, the argument is not that the author necessarily thinks that this issue is only relevant for cognitive enhancers. Rather, the goal here is to indicate that this seems to be the suggestion coming out of the overall literature.

issues regarding fairness are threatened by cognitive enhancers, non-cognitive enhancers pose a similar threat.

Interestingly, Kipke (2013) highlights one specific area where cognitive enhancement might warrant special focus: academic exams. Here, as opposed to competition for professional opportunities where emotional and social skills play an important role, what is tested appears to be mainly propositional knowledge – i.e. things that rely purely on cognitive capacities like memory, understanding, comprehension, etc. This leads Kipke to conclude that this is a unique area where cognitive enhancements specifically raise issues regarding the integrity of exam results.

I disagree. As we have seen, to the extent that Ritalin, Adderall and modafinil are non-cognitive enhancers, these drugs can be used to enhance the mood and motivation necessary for effective study, allowing users to gain a pharmacologically-induced advantage over others in an exam setting. And while exams purportedly only test our cognitive skills, in reality they also challenge our ability to remain calm under stress and to keep motivated and confident. A pharmacologically-induced sense of confidence and motivation can go a long way in helping someone in an exam setting to make the most of what they already know.

And so, it seems unjustified that cognitive enhancement in particular receives such attention in the literature: the ethical issues it raises are largely the same as those raised by non-cognitive enhancers – indeed, as I have argued, most authors are actually referring to non-cognitive enhancers when debating the ethics of PCE. To that extent, it appears that there is a kind of exceptionalism in the literature being unduly attributed to cognitive enhancement drugs. In the same way that genetic exceptionalism holds genetic information as special, requiring a unique handling compared to other types of medical information, the trend highlighted here suggests an implicit kind of cognitive enhancement exceptionalism that is, I have argued, uncalled for.

3.4 Presumption #4: There is evidence that PCE is increasingly in use

A common introductory remark in a paper on PCE is that such enhancements are increasingly in use, especially amongst students. For instance, the often-cited paper by Greely and colleagues (2008) starts off by asserting:

Today, on university campuses around the world, students are striking deals to buy and sell prescription drugs such as Adderall and Ritalin — not to get high, but to get

higher grades, to provide an edge over their fellow students or to increase in some measurable way their capacity for learning.

Similarly, Cakic (2009) notes the (then) lack of ethical examination of “academic doping”, despite “the widespread non-medical use of psychostimulants such as methylphenidate across universities for the purposes of enhancing concentration”. Likewise Turner and Sahakian (2008) ask how society should react given “the growing evidence, particularly from the United States, that pharmaceuticals are being both prescribed and illegally consumed by university students to maintain supernormal levels of concentration in the run-up to exams, with the suggestion that this trend will eventually encompass younger children.”

This starting premise of such discourses suggests a certain urgency to this debate, reflecting what looks like an emerging epidemic that is being largely ignored, potentially at society’s peril. But what is the actual evidence for this alleged trend? Here I argue that the studies and surveys regularly cited in these papers are either unreliable, or do not actually reflect what the authors are claiming.

Consider the two pieces of evidence cited by Greely et al. (2008) for the prevalence of cognitive enhancement use on US campuses. The first one surveyed stimulant use in 119 US colleges, reporting prevalence rates between 0% and 25% (McCabe, Teter, Boyd, Knight, & Wechsler, 2005). While the prevalence of 25% was found in only one out of the 119 colleges surveyed, with only 3 colleges reporting rates over 15% – the median estimate being only 3% – Greely and colleagues positioned the findings to conclude: “On some campuses, up to 25% of students had used [stimulants] in the past year”. While technically true, this rendering of the survey results is also misleading and sensationalist.

The second piece of evidence, widely used in other papers to bolster this fourth presumption, is an online poll of 1400 *Nature* readers that found one in five respondents claimed they had used drugs for non-medical reasons to stimulate their focus, concentration or memory (Maher, 2008). Needless to say, an online poll of a science journal’s readers is not an accurate representation of such drug use in a population, not to mention the credibility of online polls as a research method more generally.

Another regularly cited survey used to support this presumption found that 16.6% of 283 students in one US college (MCLA) had used methylphenidate for non-medical purposes (Babcock & Byrne, 2000).

Bush (2006) cites this finding as part of the evidence for healthy individuals using stimulants to maximize their alertness and maintain extended periods of wakefulness. Warren and colleagues

(2009) cite it as evidence that 16% of students at *some* US universities take prescription medication as study aids. Other authors cite the survey to the same end (Farah et al., 2004; Koelch, Schnoor, & Fegert, 2008)

All this despite the 16% figure being based on student responses inside *one* university answering this question: “Have you ever taken Ritalin for fun (non-medical purposes)?” Keeping in mind the question refers to *ever* taking Ritalin, as opposed to regularly using it, the survey is also only concerned with the recreational (“fun”) use of Ritalin, not cognitive enhancement. As Lucke and colleagues (2011) point out, the one mention of Ritalin being used as a “study aid” is in the report’s discussion: “Personal communications with students at MCLA suggest that methylphenidate is sometimes used as a study aid for ‘pulling all-nighters’” (Babcock & Byrne, 2000). This does not justify using the 16% figure as being indicative in any way of cognitive enhancement use among students in the US generally, let alone one that the trend is increasing.

If anything, and if we are to be as equally selective about the evidence we cite, the trend with methylphenidate at least appears to be falling, with the number of college students in the US admitting to using Ritalin steadily declining from 5.7% in 2002, to 3.9% in 2006 (Kapner, 2008). As for how often stimulants are used, a survey of 3639 undergraduates at one US university found that while 6% of students used prescription stimulants nonmedically in the last year, only 1% had done so on more than 10 or more occasions in that year, and only 2% had done so in the last month (Teter, Falone, Cranford, Boyd, & McCabe, 2010). This suggests that use of these drugs is sporadic and infrequent.

It is worth stressing that such surveys rarely ask about the use of these drugs for cognitive enhancement purposes specifically. However, more recent surveys in Europe that specifically looked at the use of these drugs for cognitive enhancement found far more modest use. In the case of Germany, life time prevalence use of methylphenidate and amphetamines for cognitive enhancement in pupils was 1.55 % and in students 0.78%, with last year and last month prevalence rates being significantly lower (Franke et al., 2010). Three large studies of German university teachers also showed a very low prevalence rate for putative PCE use (Wiegel, Sattler, Göritz, & Diewald, 2015). In the case of university students in the UK and Ireland, a majority of those survey were unaware and/or uninterested in PCEs (Singh et al., 2014). For methylphenidate and Adderall, current users and regular past users made up less than 1% of the study sample, respectively. Modafinil had a higher rate of current users at 3.9%. Interestingly, the survey suggests that students may be quite resilient to PCE, given the contrast between the level of awareness and interest in PCE (about one third of those sampled), and the comparably

far lower levels of consistent use of PCE. While this contrast may be due to the lack of availability of putative PCEs, it is also possibly a reflection of the fact that one time users find that these drugs don't make any remarkable difference to their academic performance, and hence do not pursue their use.

Despite this, it may be argued that it is still worth discussing the implications of the possibility that this trend will increase with time. It is, after all, a new trend and it would be reasonable to expect it to grow as these drugs become more effective and/or readily available. But it is important at this stage to position this trend within its historical context.

One of the problems with relatively new terms like cognitive and neuro-enhancement is that they imply the emergence of a new phenomenon. But this is not the case. To a large extent, the bioethical debate on cognitive enhancement continues to ignore the fact that many stimulant drugs that are now illegal were previously developed for therapeutic purposes and were later used by healthy individuals for enhancement purposes (Lucke et al., 2011). For instance, the early amphetamine Benzedrine – which, like modafinil was initially approved as a treatment for narcolepsy in 1937 – was used by university students to “aid the time-honoured practice of last minute ‘cramming’ for exams” (Rasmussen, 2008). The year before its approval for narcolepsy a study published in *The Lancet* reported an increase of approximately 8% on intelligence test scores among patients administered the drug (Sargant & Blackburn, 1936). As amphetamines became more widely available and began to be sold as mild antidepressants, they were also advertised as enhancing mental performance. By the late 1960s, just before amphetamine distribution became legally restricted, the estimated non-medical use of stimulants was comparable to 2002 (Rasmussen, 2008).

Interestingly, later studies showed that any perceived improvement in cognitive ability was due to these earlier stimulants' capacity to increase confidence and drivenness (Rasmussen, 2008). In that sense, they may be all the more similar to the drugs currently under discussion, which, as I have argued, appear to impact mood and motivation rather than cognitive performance amongst the healthy.

The point here is not to suggest that therefore there is nothing new to discuss. But, given the poor data at hand, and given that we have reason to believe this is not a new trend diverging from the historical abuse of stimulants, the urgency entailed by the presumption of mass and increasing use of these drugs certainly needs to be tempered.

4. The cost of hype

I have argued that the four common presumptions underlying much of the debate on PCE do not really hold up to scrutiny. The evidence for cognitive enhancing drugs boosting cognition amongst healthy users is, at best, inconsistent. Moreover, it is not clear that it is possible to pharmacologically boost cognition, without impacting mood, and vice versa. Given this, and given that, as I have argued, cognitive and non-cognitive enhancement raise the same ethical issues despite being separately dealt with in the literature, the specific focus on PCE is difficult to justify and is indeed confusing. Finally, it's not clear at all that these drugs are being used in large or increasing numbers.

It may be cautioned that these conclusions risk encouraging readers to dismiss the topic as a negligible issue for neuroethics and drug policy (Farah, 2011). However, exposing these presumptions as unwarranted is not to say that there is nothing to debate. On the one hand, how emerging technologies might impact our perception and appreciation of things like accomplishments, authenticity, justice and risk is a discussion that is worth having without the need for the hype and confusion entailed by the four presumptions critiqued here – indeed, as we will see, this is precisely the discussion that the welfarist account guides us into having. On the other, while the evidence for the efficacy and prevalence of these drugs as examples of PCE is limited, they may still be boosting performance either by inducing moods conducive to work or study, or by enhancing motivation itself. Even under such non-cognitive characterization, this would still warrant worthwhile ethical inquiry, and has indeed recently done so (Kjærsgaard, 2015; Lyreskog & Nagel, 2015).

As things stand, however, what these conclusions suggest is that the literature suffers from poor conceptual clarity along with a clear misalignment between empirical research findings and the bioethical debate. It is a misalignment that threatens the credibility of bioethics as an approach to ethical issues related to advances in biology and medicine, but more importantly it poses potential harms to the broader public due to the current relationship between the academic literature and media reporting on this topic.

The evidence for that relationship is compelling. A study of media reporting on “neuroenhancement” analysed 142 newspaper articles between 2008 and 2011 and found that 94% portrayed neuroenhancement as common, increasingly in use, or both (Partridge, Bell, Lucke, Yeates, & Hall, 2011). And while 95% mentioned at least one possible benefit of these substances, only 58% mentioned any possible risk or side effects. More to the point, the primary

sources for these claims were based on overenthusiastic reports by bioethicists and neuroscientists in the academic literature.

There is also some evidence of circular reporting, whereby the media references academic papers and academic papers reference the media. To illustrate an instance of the latter trend, a paper published in the *International Journal of Neuropsychopharmacology* cites an informal poll that was published in Cambridge University's student newspaper *Varsity* as evidence that PCE use is widespread (Mohamed & Sahakian, 2012). And yet it is precisely such papers that claim "an unprecedented rise in the use of PCEs" that arguably feed media reports.

Given this media-academia relationship, bioethicists may unwittingly be advertising the use of these drugs as part of a considerable, seemingly inevitable trend, thereby potentially pressuring the public and especially students to use them by increasing a sense of a permanent competitive struggle (G. Wagner, 2013). Indeed, this sentiment was noted in a study on stakeholder perspectives on media coverage of cognitive enhancement (Forlini & Racine, 2012). As one student put it, commenting on media coverage of the non-medical use of methylphenidate: "I don't know why but it made me want to try Ritalin."

What this suggests, again based on the relationship between media reports and claims in the academic literature, is that bioethics may end up being complicit in fostering the very trend it is meant to be merely debating. The overall outcome risks increasing and normalizing the use of these substances despite a lack of evidence for their efficacy and safety.

Altogether, this trend may increase the risk of public policy being poorly informed (Partridge et al., 2011). The literature, in conjunction with media reporting on the topic, unduly pressures society to take a premature stance: if these drugs work and they are in increasing use, we are required to resolve whatever ambivalence we feel towards these interventions and arrive at some regulatory conclusion. And there clearly is a sense of ambivalence here (Levy, 2014; Parens, 2005). The fact that there is such prodigious debate on how to think and respond to PCE reflects some degree of feeling conflicted about these interventions – otherwise, they would not be deemed worthy of so much debate. Yet we know that when people are conflicted, they seek to resolve this dissonance even at the cost of highly motivated reasoning (Cooper, 2007; Levy, 2014). Acknowledging the fact that these drugs promise very little to no improvement among healthy non-sleep deprived individuals can go a long way in encouraging a less polarized debate based on a more grounded assessment of the actual evidence.

5. A positive shift

These concerns serve as arguments for a more critical approach rooted in better empirical, conceptual and ethical research. To an extent, there are an increasing number of such calls in the literature. For instance, Mohamed (2014a) calls for larger surveys on diverse population to determine the motives and attitudes behind stimulant misuse, as well as greater dialogue between researchers, clinicians, parents, teachers and regulatory bodies in the evaluation of putative PCEs. Pustrovrh and Mali (2013) similarly stress the need for more empirical investigations on user populations in different sociocultural contexts as well as societal debates on how to address the issue. Racine and colleagues (2014) highlight broader sociological and normative presumptions in the associated literature and call for a greater awareness and explicit acknowledgement of them and how they relate to speculation on this topic.

These calls reflect a positive shift in the literature, and there is some evidence that empirical investigations are responding to such calls. For instance, recent surveys have analysed the various drivers behind a willingness to use PCE and identified several factors that increased the willingness to use these substances (Sattler et al., 2014; Wiegel et al., 2015). Amongst those were greater inclinations to use PCE when their effect was considered powerful, when more of one's peers used them, and when one suffers from academic procrastination. Newer surveys, especially in Europe and Australia, now also ask about the use of these drugs for the purpose of cognitive enhancement specifically, as opposed to their off-label use more generally (Franke, et al., 2010; Mazanov, Dunn, Connor, & Fielding, 2013; Singh et al., 2014).

6. Getting clearer about cognitive enhancement

These welcome additions to the literature mark a more fruitful approach to the topic that will hopefully prove useful in helping formulate appropriate policy responses to the prospect of PCE. In the meantime, it is clear that despite over a decade of intensive debate on this topic, the conceptual groundwork for discussing cognitive enhancement remains muddled – something evidenced by the fact that, as I've argued here, the PCE literature does not appear to be discussing drugs that specifically impact cognitive function to begin with.

Getting clearer about the nature of cognitive enhancement will be key to appropriately interpreting the implications of future empirical findings. As suggested above, aside from their

potential impact on mood or motivation, these drugs may be better characterized as “performance maintainers”, rather than enhancers in the functionalist, transmedical sense. But if these drugs work by impairing the process of fatigue, allowing one to maintain one’s otherwise standard cognitive performance over longer periods than normal, what does that say about the nature of enhancement? Might they nevertheless be a form of functionalist enhancement – i.e. the enhancement of how long one can utilize one’s unenhanced cognition for? To rephrase, if a certain mood is enhanced so that one can use their otherwise standard cognitive capacities, but for hours on end without fatigue, might one then argue that this would qualify as an example of cognitive enhancement, despite cognition not technically being boosted? Similarly, if one remains highly focused on a task because of the effects of a substance, is this due to a boost in attention or motivation? How might it be possible to tease these two apart? These questions point to the need for more thought on the nature of the interrelation between cognition, mood, and motivation, and more broadly on how enhancement and performance maintenance should relate.

For the welfarist account, with its lack of concern about functioning *per se*, these are not pressing questions, and hence I do not attempt to answer them here. In fact, according to the welfarist account, even if these interventions are ultimately deemed mere performance maintainers, they may still count as enhancements in the welfarist sense (presuming they do not come with certain other costs that may outweigh their advantages, such as addiction or some other trade-off). I move onto this topic in the next section. First, it is worth emphasizing that, as things stand in the literature, the term cognitive enhancement mistakenly implies that certain substances can work to enhance cognitive capacities generally and without trade-offs (Racine et al., 2014), and in that sense it plays right into the very hype highlighted here. Moreover, the overlap between cognitive and non-cognitive enhancement makes any reliance on this term too narrow and restrictive. It may be that we ought to do away with the term cognitive enhancement in this context entirely.

However, while “neuroenhancement” is increasingly employed to encompass both domains of our psychology, it is also potentially misleading in that it suggests our aim is to improve the functioning of neurons rather than the associated mental or psychological properties – which is presumably the actual aim behind such enhancements. In reality, it may be that we can best enhance certain aspects of our psychological properties by actually diminishing the functioning of certain neurons or areas in the brain (Earp et al., 2014). Such a term would also have difficulty accommodating the implications of the extended mind thesis, which I argued in chapter three may entail enhancing a person’s cognition by altering the environment, not their neurons.

On the other hand, terms like “mental enhancement” and “psycho-enhancement” – which perhaps better capture what is actually being referred to when we discuss such interventions – are rarely used, possibly because the connotations of words like “mental” and “psycho” may comically suggest boosting insanity in some way. Perhaps “mind enhancement” or “psychological enhancement”, while the latter is arguably overlong, offer a more appropriate rendering of what these interventions refer to and aim to achieve. For the remainder of this chapter, where I link this back to the welfarist account, I will settle for psychological enhancement, which coheres well with the account’s specific referral to changes to one’s psychology.

Such problems of nomenclature aside, what is hopefully clear at this stage is that a better, more nuanced debate will require more precision about what exactly is being referred to when discussing enhancements of this kind. That, along with an appropriately sceptical appreciation of the evidence on efficacy and prevalence, can move things forward more productively.

7. Are smart drugs welfarist enhancements?

Recall two of the central conclusions I reached regarding the welfarist account above: Firstly, given the evidence for trade-offs and impaired performance in some cases, these drugs may in fact cause disabling changes to some individuals’ psychologies. In that way, putative cognitive enhancers in the functionalist sense may not be enhancements at all in the welfarist sense. Secondly, the welfarist account’s lack of focus on functioning means it can sidestep some of the confusions arising from the interrelated nature of cognitive and non-cognitive brain states. Given the lack of clarity about which of those is primarily impacted by these drugs, but also given the inseparability of the two, we can ask the broader question of whether these drugs are psychological enhancements more generally.

In the previous section, I briefly mentioned that, even if these drugs are better understood as performance maintainers, they may still count as enhancers in the welfarist sense – that is, they may still tend to improve the well-being of their users by allowing them to maintain the use of their standard cognitive abilities for longer periods of time. In fact, even if these drugs turned out to have no measurable impact on cognitive function at all, but merely made their users feel more confident, they may still be welfarist enhancers (so long as that confidence tends to improve their well-being).

In this section, I explore the implications of these observations, and more specifically, I demonstrate how a welfarist approach to thinking about these drugs would unfold. To do this, I show how the account can be modified for a legal or political context. Crucially, I argue such a welfarist approach responds to all ethical issues raised about enhancements that are rooted in concerns about their prudential impact – that is, their impact on users, particularly in the context of a society developing them and making them accessible. This approach, I will argue, exhausts all significant ethical concerns except one: non-prudential, moral questions related to justice, fairness and equality, which is the subject of the remaining two chapters.

As we have seen, a functionalist, transmedical understanding of cognitive enhancement is only concerned with whether these drugs augment cognition in cognitively normal individuals. A welfarist approach is concerned with a number of broader factors.

The first relevant question for the welfarist about enhancement ought to be: What is the relationship between improved cognitive performance (regardless of whether this is achieved by altering cognitive or non-cognitive functions) and one's chances of leading a good life? We can rephrase this question more broadly: What is the relationship between cognitive performance and well-being? This is, admittedly, a dauntingly complex question that is burdened by, both, philosophical disagreements (e.g. what is well-being?) and empirical limitations (what is the evidence for the nature of this relationship?). It is, nevertheless, the starting point of the welfarist, though it is not a question that requires revisiting each and every time we think about psychological enhancement: once we have a good understanding of this relationship, broadly construed, it will occupy less of our deliberations as we consider different cognitive interventions.

Despite its difficulty, there are uncontroversial conclusions we can reach in attempting to answer it. Recall in chapter two that we can reach an overlapping consensus about what constitutes a good life and the sorts of things that can contribute to well-being. Employing this approach here, where a diversity of views is likely to be in play, will be crucial: a plausibly good life will include pleasure broadly-defined, desire satisfaction and a host of uncontroversial goods such as autonomy, knowledge and friendship. It need not concern us which of these are intrinsic goods and which are instrumental: what matters is that these likely play an important role in a good life in most circumstances.

Similarly, we can reach uncontroversial conclusions about the importance of cognitive ability to well-being. Firstly, presuming one must be sentient in order to have a well-being, then some cognitive ability is necessary for a person to have a well-being – hence, there is no doubt that

cognition is fundamental to well-being. Moreover, it is uncontroversial that most people want and enjoy the goods associated with greater cognitive ability: knowledge, being discerning, thinking clearly, having a good memory – these are things that most rational people are likely to want and without which autonomy is not possible. Of course, to what degree people may want these abilities is where the difficulty arises. For instance, there is no evidence that higher cognitive abilities increase or decrease happiness (Gow et al., 2005; Hartog & Oosterbeek, 1998; Sigelman, 1981). On the other hand, intelligence, broadly construed, appears to have a prominent effect on a range of socio-economic life outcomes: higher intelligence correlates with increased income (Gottfredson, 2003; Rowe, Vesterdal, & Rodgers, 1998; Zagorsky, 2007), improved health and reduced mortality (Batty, Deary, & Gottfredson, 2007; Gottfredson & Deary, 2004; Whalley & Deary, 2001), and increased life opportunities more generally (Gottfredson, 2011; Herrnstein & Murray, 1994; Murray, 2002)

Such indicators give us reason to conclude that, at least to a certain extent, cognitive ability plays a crucial role in one's chances of leading a good life. Ideally, we would want more information: which types of cognitive abilities specifically? And to what degree? Is more cognitive ability always advantageous, or does it begin to diminish our chances of leading a good life after a certain threshold? Also, do the above findings hold if everyone or most people improve their cognition beyond the current average? Answers to these questions can greatly sharpen how we view the relationship between improved cognitive performance and well-being.

Nevertheless, while we await empirical answers to these questions, we can tentatively conclude that, at the very least, low cognition may reduce the chances of leading a good life and, therefore, we at least have a *prima facie* reason to increase cognition when possible.

At this point, the welfarist about enhancement would move to a more specific question: do the drugs under discussion in this chapter improve cognitive performance? That is, the welfarist approach would now attempt answering the sole question of concern for a functionalist account of enhancement. As I have argued, it is not clear from the evidence whether they do or not. However, even if further evidence shows that they do (or some new intervention arises that does), the welfarist account would then raise a more specific question, namely: do smart drugs improve cognition *in a way* that tends to increase the chances of leading a good life? That is, the fact that they merely improve cognition does not entail they tend to improve well-being, even if improved cognition generally also improves well-being. This is because the way they improve cognition may come with some other trade-off: perhaps they cause addiction, disrupt sleep or diminish other, perhaps similarly important psychological capacities, as some of the studies cited

above noted. Understanding what these trade-offs are, and how they are likely to impact well-being overall, will be central to establishing whether these drugs are welfarist enhancements.

However, these are not the only trade-offs to consider in relation to leading a good life: it may be that, as some have argued, smart drugs improve cognitive performance, but do so at the cost of reducing the value of our accomplishments or making us less admirable (e.g. Bostrom & Sandberg; Goodman). These concerns, if valid, may mean that these drugs in fact tend to *reduce* our chances of leading a good life overall. In that way, these drugs may be disabling either because they do not improve our cognitive performance overall, or because they reduce value in our lives in some other way that again, overall, outweighs their advantageous effects on our psychologies.

To be clear, I say *overall* because even if improved cognition tends to improve well-being, that may be outweighed by the concurrent diminishment to well-being from the loss of authenticity or meaningful accomplishments.

However, for the welfarist, such concerns, in and of themselves, would not mean these drugs are disabling. Here, the discussion will revert back to thinking about what constitutes a good life. For instance, let us assume using smart drugs entails leading an inauthentic life. Is an inauthentic life necessarily a bad life? Might improved cognition that comes at the cost of inauthenticity still have an overall positive impact on well-being? Often, in-principle arguments against enhancement (in the functionalist sense) seem to presume that if they are correct (e.g. if enhancement entails a loss of authenticity), then it must be that their use is wrong or at least bad for the individual who uses it. However, this is a presumption about the nature of a good life: one that says that, no matter the benefits, those are always outweighed by a loss of authenticity. Clearly, this would be a controversial claim and the welfarist about enhancement should not presume its veracity. Here, the debate would shift about the relationship between authenticity and well-being.

In this instance, I focused on authenticity, but we can see how the account would be concerned with other factors. For instance, we can see the concerns raised about smart drugs most clearly in the case of health and safety: if, say, modafinil is not safe, or not safe enough, whereby its use would likely disadvantage the user more than advantage them, then it is not an enhancement for the welfarist.

The same applies to other concerns, and here I am referring to functionalist enhancements more generally. Mehlman (2012) worries functionalist enhancements may diminish central features

of our identity as human beings. For instance, it may be that suffering, ageing and dying, along with our attempts to deal with these realities, is what makes our lives valuable (Parens, 1995). If true, interventions that allow us to sidestep these realities would rob our lives of value. Again, if such a disadvantage outweighs whatever other benefits the intervention confers, it is not an enhancement in the welfarist sense. In fact, it would be a source of disability.

Similarly, consider Habermas's (2003) concern that genetic manipulations may rob offspring of their authentic abilities and their autonomy to create themselves as others normally do. What is the implication here? It would be bad *for* these children; their lives would go worse overall. Again, if true, then such genetic interventions would not be enhancing – they would be disabling.

Note that even if these concerns are valid, these interventions may still be enhancements if other advantages of the intervention outweigh these supposed disadvantages. For the critics, their framing of these concerns suggests they think these disadvantages are overwhelming – no amount of other advantages could outweigh them. Nevertheless, in evaluating these arguments, the welfarist must also consider how they weigh up against other advantages. It may be that, upon further thought, interventions that render users inauthentic or addicted, that rob children of autonomy, or that diminish the value of our lives, are still good for a certain person overall, given their particular circumstances. If so, they would be enhancements.

At this stage, we can begin to see how the account provides a framework for thinking about an intervention like so-called smart drugs. The authors of the above objections may not have been motivated by welfarist considerations, but the welfarist account allows us to cash out those concerns through the language of well-being, and assess it on those terms. The effect is that the concerns are immediately clearer: we may disagree about whether autonomy or authenticity have intrinsic or instrumental value, but we can agree that they play an important role in our ability to lead good, meaningful lives, based on an overlapping consensus of what those are. But we also recognize that other things play such a role, such as success and intelligence, and so we must weigh those against each other before we can reach conclusions about their overall impact on a person's well-being. In that way, this links back to my primary goal of having an account that can help us think more clearly about enhancement, and in this case, one that can allow bio-conservatives to articulate their arguments in ways that avoid, as I argued in the introduction to this thesis, being shut down as irrelevant or confused.

With that in mind, the welfarist account certainly does not side-step ethical issues (as some of its critics I noted in chapter two have argued), but it can help us deal with those issues in quite a systematic way. Note that these are ethical issues in the context of a society developing and

marketing an intervention that changes the body or mind. The welfarist account compels us to deal with these issues before we can deem any such intervention an enhancer. Crucially, such an approach actively responds to the concerns about hype highlighted above. As I argued in section six above, the mere label of “enhancement” can give prospective users a false understanding of what the intervention does. Better to have an idea of how such interventions tends to impact their users’ well-being (rather than merely their functioning) before we call them enhancements.

7.1 The welfarist account and group enhancement

We can articulate the questions the welfarist about enhancement would raise more abstractly as a heuristic for thinking about enhancement. Any time an intervention y promises to change some bodily or mental state x , the welfarist account raises questions about its impact on well-being:

- What is the relationship between x and well-being?
- How will changing x through y tend to impact well-being?

Here, x could stand for a change in cognition or mood, or more specifically, abstract thinking or cheerfulness; it could stand for longevity, autonomy, social skills, and so on. Y in turn could refer to any bodily or psychological interventions (genetic, pharmaceutical, surgical, and so on.)

However, one challenge facing the welfarist account at this stage is its inability to give consistent answers to whether the drugs discussed here are enhancing or not: it will depend on the individual using them and their specific circumstances. For instance, in the case of methylphenidate, whether the person taking it has ADHD or not may play a big role in whether the welfarist account labels its use enhancing or disabling (or neither).

The problem with this relates to regulatory issues: the welfarist account is limited to telling us whether these drugs are enhancers or not on a case by case basis. In chapter two I argued this is not necessarily problematic: we often also cannot tell whether a treatment will actually be therapeutic except on a case by case basis, and only often retrospectively. However, for the purposes of regulation and policy, it is difficult to see how such an individualist approach will be

helpful. To that extent, the welfarist account arguably lacks the ability to answer the broader question of how an intervention may affect a given population in general.

Kahane and Savulescu (2009) are aware of this limitation in their discussion of the welfarist account of disability, noting that what disadvantages one person may not disadvantage another. What about the effect of interventions on a population of users? Here, they suggest we can alter the level of magnification to something more pertinent. Rather than speaking of changes to the body or mind of a person, we can speak of changes to the bodies or minds of most people from a given population. Kahane and Savulescu adopt this approach for defining disability in political and legal-contexts where our conclusions (in the form of laws and regulations) need to be generalizable to populations rather than particular individuals in particular circumstances.

For such purposes, we can characterize enhancements as:

Any changes in the biology or psychology of people sharing relevant circumstances that increases their chances of leading a good life.

These shared circumstances can refer to a particular demographic with identifiable and relevant circumstances (e.g. those suffering from ADHD). How exactly an increase in well-being for a group should be measured is a further question: are we looking for a likely increase in aggregate well-being of the group, an increase in the average well-being of the group, or perhaps an increase for the worst-off in the group? I do not want to explore this issue at this stage as it occupies much of the next chapter. For now, note how this re-formulation of the account can help us answer whether the drugs discussed here are enhancements in the welfarist sense. We already know that for a certain group of people – for instance, those with ADHD or sleep shift disorder – these drugs do indeed improve their cognitive abilities or regulate their sleep patterns. Presuming such improvements tend to contribute to their well-being, they would be enhancements for this group of people. On the other hand, we can be fairly sure these drugs would not contribute to the well-being of infants as a group, given the increased risk in side effects and the fact that infants do not generally benefit from improved cognitive performance at that age. Thus, we may reach different conclusions depending on the group of people we have in mind – and this in turn will be relevant to policy-makers.

Importantly, a crucial point arises once we apply the welfarist account to groups of people rather than individuals that relates to positional goods. These are goods that are valuable only to the extent that few others have access them (Brighouse and Swift 2006; DeGrazia 2012). That is, they confer a competitive advantage over others – if everyone had them, no such competitive

advantage would be gained. With that in mind, let us assume that a particular student taking modafinil will be enhanced: modafinil, with its positive effects on their grades, will tend to increase his or her chances of leading a good life. However, once we ask whether modafinil is an enhancer for university students as a category, we may reach a different conclusion. At the level of groups, we must ask whether the advantages it confers to students are primarily positional.⁵ If they are, then modafinil's impact on well-being is dependent on users within that group remaining a minority. Here, modafinil may be an enhancer for the individual student. However, it is not an enhancer for the category "students". If all students take modafinil, and if its effects are primarily positional, then its advantageous effects are significantly reduced (while the disadvantageous effects, e.g. side-effects, are not). Hence, it may be enhancing for a person taking it where their circumstances entail they are amongst a minority taking it, but for the category of students, where the relevant circumstances include all students potentially taking it, its use may be self-defeating.

In contrast to enhancers that act primarily as positional goods, some enhancers may only be beneficial to the extent that they are used by a majority. For instance, certain benefits of cognitive enhancement may emerge only through the network effects of numerous people using them: only then might, say, new forms of productivity or perhaps more fulfilling kinds of relationships emerge (Buchanan, 2011). Here, for an individual using this intervention on their own or among a minority, it is not an enhancer or at least not a significant one. However, once a large enough group is using it, it becomes a significant enhancement.

What this shows is that, at the level of groups, the welfarist account is sensitive to whether an intervention confers a primarily positional advantage or not, and similarly whether the intervention has an enhancing network effect. If it is primarily a positional good, while it may still be an enhancer for an individual, from a policy perspective focused on regulating its use for groups of people as a whole, it is not. In such a case, its use in that group would mean it confers little or no advantage, and its use may be overall disabling. In contrast, if the enhancement primarily works through a network effect, while it may not be an enhancer for an individual, from a policy perspective focused on regulating its use for groups of people, it is.

Note that in all the preceding discussion of this section, we have been discussing interventions in the context of whether they are good or bad *for* individuals or groups of people – this is

⁵ I say primarily because few goods are purely positional. For instance, even if high intelligence confers a primarily positional advantage (by granting one access to highly competitive positions and salaries), it also certainly has what might be called independent value: the understanding of the world that intelligence brings is just as valuable even if many others have that same understanding (Agar, 2008).

because we are discussing whether they are welfarist enhancements or not. As I have argued, doing this goes a long way in responding to ethical concerns about a society developing and making accessible innovations like smart drugs: to consider them enhancers, the welfarist account provides a framework for thinking about how they are likely to impact a person or group's well-being. However, this does not exhaust addressing all ethical concerns: other non-prudential issues remain, or specifically one broad issue remains. The fact that an intervention may be of prudential value to someone or a group does not mean its use or development is also fair or just. For instance, the fact that modafinil may be advantageous does not mean that we should have developed it or made it accessible, nor that we should further develop more or better drugs like it. This is because the welfarist account may consider an intervention an enhancement even if its development or use is deeply unfair or promotes injustice – an enhancement that is unaffordable to a majority is still an enhancement even if that lack of access leads to a highly unequal or oppressive society, or even if its development comes at the cost of neglecting other, more vital enhancements. In other words, the welfarist account, in and of itself, does not require us to settle all ethical issues raised by enhancement before we deem an intervention to be one.

This suggests that an ethics of welfarist enhancement would be primarily restricted to issues of justice. When is it right to improve people's lives at the cost of others? What issues arise when we consider investing in particular types of enhancements? How can we stop enhancements expanding the ability divide, and should we endeavour to do so?

I have argued the account's focus on well-being can help us grapple with the concept of enhancements. In remaining two chapters, I argue that a similar focus on well-being can also resolve these remaining questions. Specifically, in the next chapter, I argue that striving to maximize overall well-being impartially through enhancement, while blind to distribution, nevertheless adequately responds to these concerns.

Chapter 6

Enhancements and Maximizing Well-Being Impartially

So far, this thesis has been an extended examination of the welfarist account of enhancement: from what that account is, to the extended mind thesis's implications for it, to how it might be reconciled with the notion of moral enhancement, and finally, to the deployment of the account in the context of the debate on smart drugs. These have been concerned with answering this question: What is the welfarist account of enhancement, and what are its full implications?

It is now time to move on to the second question I highlighted in the introduction to this thesis: On what basis should enhancements be developed and distributed? As I noted at the end of the last chapter, the fact that a bodily or mental intervention tends to improve the well-being of a person or a group of people means it is an enhancement, but that does not therefore also mean that developing or distributing that enhancement is permissible. There may be a number of reasons why a welfarist enhancer should not be developed or distributed: perhaps resources should go to developing or distributing some other good, or perhaps it should be developed but its permissibility will depend on how it is distributed (will everyone have access to it, and if not, why not?). These are issues broadly related to distributive justice and these two remaining chapters focus on the basis through which we should approach this issue when it comes to enhancement. Here, in line with defending a welfarist understanding of enhancement, I wish to defend a welfarist principle for distributive issues related to enhancements. Specifically, I wish to defend utilitarianism – that moral theory Bernard Williams dismissed in 1973 saying, “The day cannot be too far off in which we hear no more of it” (Smart & Williams, 1973).

In the introduction to this thesis, I mentioned that, in the same way well-being can act as a common currency for cashing out other values in helping us think more clearly about enhancement, the welfare-maximizing approach of utilitarianism can similarly help resolve questions about the development and distribution of enhancements. To be clear, the welfarist account is compatible with any ethical theory as it pertains to distribution: one may be a Kantian or Aristotelian welfarist about enhancement, and future research may well explore such combinations. For my purposes, I am particularly interested in arguing for the practical usefulness of a focus on the impartial maximization of well-being as a guide for developing and distributing enhancements. I believe an approach that promotes the development and distribution of welfarist enhancements that can expectably maximize overall well-being – which

is what I will argue a utilitarian approach would endorse – is a fruitful way for practical ethicists and policy-makers to think about distributive issues raised by (welfarist) enhancements.

My goal here is not to make a case for why utilitarianism is the correct ethical theory. Instead, my task is to show how it is well suited as a “metamorality” – a global moral philosophy that can adjudicate among competing moralities (Greene, 2014). Specifically, it is one that has the resources to accommodate different conceptions of the good life and fair distribution by relying on a common currency of human values: well-being. Here, I explore utilitarianism as a metamorality in the context of distributing enhancements. My argument will not be that well-being is necessarily the one true value, but that it is a pragmatic metric of value against which other values can be measured, enabling us to reach principled compromises when such values conflict (Green 2014). Specifically, I am appealing to utilitarianism as a metamorality to guide collective rather than individual action – it is within this domain of public policy that it is arguably most persuasive. As Robert Goodin (1995) put it:

The strength of utilitarianism, the problem to which it is a truly compelling solution, is as a guide to public rather than private conduct. There, virtually all its vices - all the things that make us wince in recommending it as a code of personal morality - loom instead as considerable virtues.

In that regard, I am arguing for institutional utilitarianism – the view that the best political and social institutions are those that produce the greatest total well-being (Mulgan, 2014a) – rather than act utilitarianism. However, as we will see, even as a guide to public conduct and the institutions responsible for distribution, utilitarianism still is accused of a number of vices. Before I respond to these, I first elaborate on the basic components that constitute utilitarianism in section one, arguing broadly for the compatibility of those components with most plausible ethical theories – an approach that links to my aim of seeking overlapping consensus about theoretical and conceptual disagreements where possible. In section two, I move on to explore some preliminary implications that utilitarianism has for the distribution of enhancements, which also sets up one of the objections against utilitarianism – namely, that it may lead to an oppressive or unfair distribution of enhancements that may violate basic rights. To understand this objection fully, in section three I first elaborate on the principle of diminishing marginal utility along with what has been dubbed the wealthitarian fallacy (Greene & Baron, 2001). It is at this stage that we can begin critically examining this distributive objection in section four, along with another objection that utilitarianism may neglect impaired individuals (or, in our context, healthy but unenhanced individuals). Finally, in section five I explore how the state, which plays a central role in distribution, can use utilitarianism to distribute enhancements, but argue for why it is unpragmatic to rely on states to fulfil such a role effectively.

1. Utilitarianism's building blocks

Three logically independent components constitute utilitarianism (Sen & Williams, 1982). These are:

Welfarism: The thesis that ethics is exclusively concerned with well-being.

Consequentialism: The thesis that actions, motives, rules, institutions and other entities should be chosen solely on the basis of the states of affairs they produce.

Sum-ranking: The thesis that the value of a state of affairs is determined by nothing other than the total quantity of value it contains (Eggleston, 2014).

While I elaborate on these components below, it is worth first putting these together to summarise what they entail utilitarianism is about: *maximizing well-being impartially*. The impartiality arises from how we weigh up interests, which I take to refer to whatever it is that contributes to well-being when fulfilled or not conflicted with (these may be desires, informed desires, pleasure, avoiding pain, success, and so on). Here, if our goal is to maximize well-being, then the only way to do that is to weigh equal interests equally (or similar interests similarly). If two individuals have the same interest (for instance, not being in pain), and if we seek to maximize well-being, then we must give these interests the same weight in considering how to respond to them (all else being equal). If we do not, it would mean that the way we are attempting to maximize well-being is based on false calculation – that is, we would not actually be maximizing well-being. To do that, we must be impartial in weighing interests.

This observation leads to the egalitarian foundation of utilitarianism known as *equal consideration of interests* (Singer, 2011). DeGrazia (2008) explains it as such: “To give equal consideration to A and B is to hold that we should grant roughly equal moral weight or importance to A’s and B’s (prudentially) comparable interests.” It is an idea that can be traced back to Jeremy Bentham: “Each to count for one, and none for more than one” (attributed to him by Mill, 1863). To be sure, this egalitarianism about interests is not because of a commitment to egalitarianism itself, but simply because in aiming to aggregate well-being honestly (which is to say, effectively) we must acknowledge that, as Singer (2011) puts it, “an interest is an interest, whoever’s interest it is.”

1.1 Welfarism

Welfarism entails a commitment to explaining and justifying the goodness or badness of anything based solely on its contribution to wellbeing-capable creatures (Crisp, 2016). For the welfarist, then, an ethical theory that claims anything other than well-being has intrinsic value, whether it is autonomy, knowledge, respect for rights, beauty, virtue or the realization of human excellence, is false. Such values may be of great importance to the welfarist, but only to the extent that they instrumentally impact well-being. Only well-being has intrinsic value.

Welfarism's appeal lies in its intuitiveness. Why else can something be deemed good but for the fact that it makes lives go better? Surely something is good precisely because it *makes* lives go better. Similarly, what is it that makes something bad if not for the fact that it makes lives go worse? This seemingly axiomatic appeal of welfarism makes it difficult to articulate a positive argument for it. Sumner (1992) conveys this sentiment well:

The idea that the unifying - and justifying - function of all of our ethical categories is ultimately to make our lives go better, or to make the world a better place, is one that I find utterly compelling. If that is not the point of the whole business of moral thinking, then I find it difficult to imagine what the point might be. What else could morality be for? And if it is not for anything - if it has no point - what claim can it have on our allegiance?

These questions Sumner raises are not easy for the non-welfarist to answer – that is, someone who believes something can be ethically good or bad without being good or bad for anyone. Moreover, whatever the non-welfarist values, the welfarist is in a good position to respond by either pointing out why such values are only instrumentally valuable – why else would respecting others be valuable if not for the fact that doing so is good *for* someone – or by pointing out that it is non-morally valuable – a beautiful world with no conscious creatures may have aesthetic value, but what happens to that world is of no consequence to ethics.

These responses arguably enable welfarism to accommodate most of our intuitions about moral value, regardless of whether those values turn out to be inherently or instrumentally valuable. That is, in the domain of public affairs at least, the welfarist can agree with what the non-welfarist values in practice: for policy purposes it does not matter whether what the non-welfarist values is intrinsically valuable or not – theoretical disagreements can be bypassed by

agreeing such things are valuable without having to agree whether they are of intrinsic value or not.

Furthermore, welfarism offers a compelling account of what endows something with moral status: we do not ordinarily consider rocks and plants as having moral status precisely because we have no reason to believe they have a well-being. Similarly, there is something compelling about ascribing more moral status to creatures with increasingly complex neurologies, to the extent that we have reason to believe there is a close relationship between neurological complexity and the ability to have a good or bad life (DeGrazia, 1996).

To my mind, while it is beyond this thesis to present a comprehensive argument for welfarism, this gives us a number of reasons to favour welfarism. Even if its exclusivity claim about well-being is wrong, it still seems highly implausible that well-being isn't good.

To that extent, it is a strong position that can accord with common sense and plausible accounts of the good, even non-welfarist ones, to the extent that these agree welfare is intrinsically good.¹

Nevertheless, identifying well-being as the sole concern of ethics does not mean much without an account of what well-being is. Is a good life merely the satisfaction of one's desires? Is it purely a function of the pleasure one attains? Or is life's goodness determined by its relation to a list of "objective" goods? In chapter two and five, I argued that disagreements about what the nature of well-being is are not problematic for the welfarist account of enhancement because these approaches tend to converge on similar items. This conclusion applies all the more here: As I state above, I am advocating utilitarianism not as the (necessarily) correct moral theory, but as the one that is likely to be most practical for political theorists thinking about distributive issues. Given that aiming to promote the good(s) advocated by one account of well-being will tend to also promote the good(s) advocated by others, philosophical disagreements about what well-being is become of little practical importance. I explore this issue further in section five.

Ultimately, in the context of enhancement, welfarism entails that when it comes to decisions to develop or make available an enhancer to a public, such decisions should be guided solely by how they are likely to impact well-being.

¹ For readers who think welfare isn't good, you may proceed with what follows as an extended hypothetical or an argument conditional on welfare being good.

1.2 Consequentialism

Having identified utilitarianism's account of value, we can now explore what it prescribes as the correct response to that value. A consequentialist account is distinguished by its claim that the correct response to an ethical theory's identified value is to promote it (Pettit, 2002).

In contrast, the non-consequentialist may claim that the correct response to a value is, at least in certain cases, to honour it – that is, to set constraints or restrictions that protect it from damage or potential damage (Pettit 2002). This entails that there can be occasions when one must not violate such restrictions, even if doing so is the only way to prevent a greater violation of that same constraint on others. This is the so-called paradox of deontology (Scheffler, 1988).

The non-consequentialist here needs to explain why we should sometimes refrain from promoting an identified value, even when doing so would reduce the number of violations of a deontological restriction. Nozick (1974) raised this issue through a compelling question: "How can a concern for the non-violation of C [i.e. some deontological constraint] lead to refusal to violate C even when this would prevent other more extensive violations of C?" A classic example of this is the trolley problem: non-consequentialism with a restriction on killing will allow for five people to be killed rather than one, even though killing one person will reduce the number killed by four. One need not be an absolutist to concur with this: moderate non-consequentialism would also place restrictions that it would be wrong for one to violate, at least in some circumstances, even though doing so would minimize violations of the same restriction overall (Shaw, 2009).

This paradox of deontology can appear illogical, and some have argued that non-consequentialists need their own model of rationality that can compete with the more intuitive "act-to-produce-the-best-consequences" model motivating consequentialists (Alexander & Moore, 2015). And while I do not claim that any such non-consequentialist model of rationality is bound to fail, it is difficult to see why anyone would pursue it, if not for a belief that consequentialism has the indefensible implications that I will discuss and offer a refutation of in section four. More importantly, "act-to-produce-the-best-consequences" has the appeal to be part of a meta-morality in the context of political decision-making: in formulating policies and regulations, setting up non-consequentialist constraints that, in some cases, may limit how well people's lives could go, will be difficult to justify. Public policy that is not rooted in trying to make lives go as well as possible, or that benefits no one while harming others, all else being equal, is not defensible.

Ultimately, consequentialism can be compatible in practice with deontological constraints: to the extent that people want to restrict violations that may damage a certain value, consequentialism can factor this into account if ignoring their wishes will not produce the best consequences. Hence, while there may be theoretical disagreements about why a certain value should be honoured rather than promoted, consequentialism offers a way to bypass these and reach agreements that are satisfactory in practice.

1.3 Sum ranking

So far, I have identified what is of value for utilitarianism (well-being) and what it prescribes as a response to that value (promote it). The remaining question is how the theory evaluates the outcome of promoting it. There are a number of ways to do that (Eggleston 2014). An egalitarian approach considers an outcome better when the distribution of well-being in it is more equal than it otherwise would have been. Prioritarianism holds the best outcome as the aggregate of total well-being while also holding that more well-being for the worst off is better than more well-being for the better off. Sufficientarianism holds the best outcome as the one where the greatest number of people have a “sufficient” degree of well-being. Finally, one can merely be aggregative about well-being, looking at the total amount of well-being summed up across all conscious creatures, with the more total well-being the better (regardless of how it is distributed).

This sum-ranking approach characterizes utilitarianism. It is also arguably raises the most objectionable implications for utilitarianism, as its lack of sensitivity to distribution implies that it could endorse violating basic rights and could lead to extreme inequality in some (at least hypothetical) scenarios.

I will consider such objections in detail in section four, but for now, it is worth briefly highlighting how each of the other aggregation functions mentioned above also face serious difficulties. For instance, egalitarianism appears to endorse “levelling down” – that is, it would approve reducing the well-being of a great many if it led to more equality (Temkin, 1993). Such a move would also conflict with welfarism, as such a monistic focus on equality would hold it as inherently valuable, possibly as the only thing that is so.

For prioritarianism, it raises the question of how to non-arbitrarily give weight to (that is, prioritize) the worse-off but while also improving the well-being of the better off in some

circumstances. At which point exactly would giving more weight to the worse off instead of, say, a relatively well-off majority, be wrong? Moreover, prioritarianism appears to give greater weight to the interests of some over the identical interests of others, and this arguably violates the basic utilitarian premise of equal consideration of interests. It also faces some of the concerns about sum-ranking to the extent that it is also aggregative.

For sufficientarianism, we are faced with the question of how to non-arbitrarily set a threshold of well-being that is deemed sufficient. It also seems to endorse neglecting people who fall far below the sufficiency line and for whom no amount of resources will get them over it, instead theoretically supporting the use of those resources to move a single individual just barely below the line to just barely over it (Arneson, 2013).

Though it is beyond my scope to explore this here, I suspect that combining elements of these approaches into hybrid theories leads to merely combining several of these defects into one theory (as in the case of prioritarianism), not to mention that doing so would be uninvitingly ad hoc.

Nevertheless, it is worth conceding that, at least in the cases of egalitarianism and sufficientarianism presented here, these are extreme or simplistic versions of them. More plausible versions would probably need to – like prioritarianism – incorporate some purely aggregative, sum-ranking element to it (Arneson, 2013). However, this means that, at least in some hypothetical scenarios, the objections against sum-ranking would similarly apply to these more modest versions of egalitarianism and sufficientarianism. If so, that is, if the implications of sum-ranking are not avoidable in more plausible versions, then why add further problematic elements to our theory, given that such attempts would not circumvent those implications? In that sense, sum-ranking comes off as the most parsimonious, with a straightforward commitment to welfarism and a non-arbitrary foundation. It is a good default candidate – all the more so as we explore my response to objections against utilitarianism below.

Ultimately, I suspect that, whichever aggregation function is selected, it will not have major implications in the real world. I believe that the conclusions utilitarianism reaches in policy contexts are likely to be broadly acceptable to moderate egalitarians, sufficientarians and prioritarians. Moreover, as I argue below, the major objections to a sum-ranking approach are based on a recurring misunderstanding of utilitarianism where utility is mistaken for wealth. I will return to this issue in section 3.1

2. Utilitarianism and enhancement

The welfarist account of enhancement is, I have argued, an approach to understanding enhancements through their prudential value. I have emphasized that the fact something is an enhancement (and therefore tends to be good *for* the person who undergoes it) does not mean that its development is permissible, or that it should be distributed, whether privately or publicly – other values, such as justice, need to factor into our ethical assessment of it. For utilitarianism, the determining factor for whether we ought to develop a certain enhancer, or how we ought to distribute an existing enhancer, is the effect on overall well-being. We can express this through two questions. If the enhancement does not yet exist but may be developed, the question is:

Will its development expectably maximize overall well-being?

If the enhancement already exists, the question is:

In what way should it be distributed in order to expectably maximize overall well-being?

Note the use of “expectably”. The difference between framing utilitarianism in terms of actual as opposed to expected outcomes is subtle but may have important implications if our goal is to ascertain the precise nature of what the “true” ethical theory is. Given our pragmatic purposes here, however, we can speak in terms of expected outcomes. In practice, there is little difference between both approaches: long-term, the only reliable way to maximize actual well-being is to act in ways that are expected to do so (Shaw 2014).

With that in mind, let us further explore these questions. Note that it is not sufficient for a utilitarian to answer these questions by merely weighing costs and benefits. In terms of deciding on whether to develop an enhancer, there must also be no alternative course of action that would expectably lead to more well-being – even if that alternative course of action is completely unrelated to enhancement. Only in such cases would its development be permissible. Similarly, when deciding on how to distribute an existing enhancer (e.g. whether to distribute it all, whether to subsidize it or tax it, or require a prescription or license for it), the course of action chosen must be the one that would expectably maximize well-being.

However, even if we determine that developing a certain enhancer or distributing it in a certain way is permissible, the utilitarian is faced with a further question: is its development or

distribution also required? For that to be the case, developing or distributing the enhancer in a certain way must result in greater expected well-being than anything else we could do. On the other hand, if the well-being expected from some other course of action would lead to an *equivalent* expected effect on well-being, there is no requirement to select one over the other: choosing either would be right according to utilitarianism. While this may sound like a theoretical point that is unlikely to be relevant to real world decision making, note that given the difficulty of estimating expected well-being, such ties between choices may be more common than we would otherwise expect (Shaw, 2014).

With these preliminaries out of the way, let us consider which sorts of enhancements utilitarianism would prioritize and which it would discourage. Let us start with the latter. Here it is worth recalling the list of six relationships that prudential and moral value can have when it comes to changes to an individual's biology or psychology (chapter four, pp. 79). Clearly interventions that are prudentially disadvantageous and that reduce moral value would be discouraged, as would interventions that leave users prudentially worse off while reducing moral value. From the perspective of maximizing overall well-being, these fail outright. Here, for utilitarianism, the combined value of prudential and moral value – that is, the impact on the user and the impact on others from the user undergoing the change – *is* the overall impact on well-being from an intervention. No doubt, determining what the prudential and moral value of an intervention is will be challenging in some cases. Here, I am assuming a society has a well-developed account of well-being based on some overlapping consensus. To the extent that such an account cannot accommodate all conceptions of a good life, and to the extent that paternalism is associated with a diminished well-being, utilitarianism would limit how much it interferes with people's personal choices when it comes to such well-being diminishing interventions. To be precise, it would limit such interference precisely to the extent that doing so would also maximize well-being.

Things get further complicated with interventions that are prudentially advantageous yet reduce moral value. An example worth considering are enhancements that act primarily as positional goods, which we encountered at the end of the last chapter. Recall I argued that, in terms of groups of people, interventions that are mostly positional goods may not be enhancements: their effect on a group who collectively makes use of it would be largely self-defeating. However, what about individuals? In their case, such interventions would be prudentially advantageous (they *are* enhancements) while at least potentially reducing moral value: they primarily work so long as others are forbidden from these same advantages, and if everyone uses them, there is little net gain to anybody (Bostrom, 2003). As such, all else being equal, the more an

enhancement is closer to being purely a positional good for an individual, the more utilitarianism would discourage its development and, if feasible, limit its distribution. On the other hand, if it is primarily a positional good with some independent value and has already been developed and is in fairly common use, utilitarianism may encourage its wide distribution. This would reduce its positional aspect (if everyone has it no one gains a competitive edge), while retaining whatever independent value it may have.

Things get similarly complicated with interventions that prudentially disadvantage users but that create moral value – that is, in the context of utilitarianism at least, this is an intervention that reduces the well-being of the user but increases overall well-being through its impact on others. For utilitarianism, these are likely to be used in the context of deterrence or punishment, where an individual is intentionally disabled in the welfarist sense (say, someone like Pido from chapter four) with a net gain for society (children are safer and other paedophiles may be deterred).

In contrast, interventions that would be prioritized by utilitarianism are those that are prudentially advantageous while also increasing moral value, and those that leave users prudentially the same but increase moral value. However, from the perspective of maximizing overall well-being, things again get complicated as we try to assess which interventions *would* create the most value overall. The more overall value (that is, the more prudential and moral value combined), the more priority or encouragement it would be given.

So far, I have been speaking in abstract terms: as usual, things get all the more convoluted when translating this into concrete examples. These, as we have seen in the previous chapter, are complicated by sometimes conflicting evidence on efficacy and safety, and a quagmire of social and personal side-effects that are often difficult to foresee. Nevertheless, it is worth noting this: given that our conception of enhancement includes vital therapies in its definition, utilitarianism would prioritize enhancers that save many lives and reduce debilitating pain². More generally, utilitarianism would prioritize those enhancements that change the bodies and minds of people in ways that tend to *most increase* their chances of leading good lives.³

Note the emphasis on *most increase*: acne and cancer treatments may both increase well-being, but the latter are likely more vital to well-being and as such would be prioritized by utilitarianism. Of course, this will depend on the number of people who suffer from cancer and

² Regardless of the account of well-being used, being alive and free of debilitating pain are preconditions for a good life, whether that be one of pleasure, desire satisfaction, and/or the attainment of various inherently valuable goods.

³ Based on my description of utilitarianism, it would not limit this scope to people, but to all conscious creatures capable of well-being. I limit it to people here given the scope of my topic.

acne and their relevant circumstances (for instance, if a society has a tradition of killing its teenagers who develop acne, utilitarianism may prioritize pre-emptive acne treatments over cancer treatments). Hence, in general, those enhancements that extend lives while reducing how painful or debilitating a condition is will top the list for utilitarianism. These are interventions that, relatively speaking, reliably contribute to over-all well-being.

Given the extent of ill health and suffering in most societies, it can be difficult to see how utilitarianism would currently find developing enhancers that add new capacities or augment current ones beyond their standard functioning permissible, let alone required. It is difficult to see how the well-being derived from such augmentations can compete with vital treatments. As such, utilitarianism may appear quite hostile to the development of transmedical enhancers.

In reality, the picture is more complex and here two points are worth highlighting. The first relates to what the expected effect of a given enhancer is. Imagine having to choose between a psychological enhancer that will treat an individual (call him Lennie) for a cognitive deficit. In simplistic terms, we can imagine it will increase his IQ from 70 to 100. Another enhancer will augment an already intelligent person's cognition (call him George) from, say, having an IQ of 120 to 150. As enhancers, it may be that Lennie will more greatly improve his chances of leading a good life compared to George. For Lennie, the change in his cognitive ability may have huge implications for things we believe are closely tied up with a good life: meaningful, intimate relationships, achievement, autonomy, and so on. In contrast, George may already be well placed to have these things in his life. In other words, enhancing Lennie may be more prudentially advantageous for him compared to George.

However, for utilitarianism this comparison is not sufficient. This is because, if George's new cognitive abilities expectably resulted in him, say, partaking in innovations and breakthroughs that have a huge benefit for a large number of people across generations, then his augmented cognition would contribute more to overall well-being than Lennie's. That is, though he may gain less prudential value from the intervention than Lennie, he will create enough moral value to outweigh the overall well-being arising from enhancing Lennie. In that way, utilitarianism may favour enhancing the already cognitively healthy rather than treating those with cognitive deficiencies, if doing the former is likely to lead to such consequences (Dunlop & Savulescu, 2015).

The other thing to note here is the issue of future generations: utilitarianism is not only concerned with present well-being, but overall well-being and hence well-being across all time and space. This concern has a number of implications for the theory, and many critics use them

to object to it (Mulgan, 2011, 2014b; Vallentyne, 1994; Van Liedekerke & Lauwers, 1997). For instance, even if the human population were to remain static at around 4.5 billion, it is estimated that another 500 trillion humans are yet to be born if we survive the next 10 million years, which is the mean lifespan of a “successful” species (Sagan, 1983). This means that, presuming we are not made extinct soon, the vast majority of human suffering and well-being is yet to come and it entails that our obligations to the future far outweigh our obligations to those who currently exist (except to the extent that future generations are contingent on the existence and relative well-being of those who currently exist). As such, utilitarianism is particularly concerned with existential risks: asteroids, nuclear wars and climate changes that threatens the species as a whole are particularly troubling for the utilitarian. To the extent that certain transmedical enhancements can protect against such existential risks, utilitarianism is likely to rate them highly in terms of their ability to help ensure future generations come to exist.

Ultimately, what is clear is that utilitarianism is not the sort of ethical theory that can make *a priori* claims about which enhancements should be developed and used and which should not be: as with the welfarist account of enhancement, all is context dependent according to utilitarianism, with the specific details of a given enhancement playing determining roles in the permissibility or requirement of an action.

With this modest conclusion reached, we can already start to see a worry arise from this approach: in the pursuit of maximizing overall well-being, utilitarianism can potentially endorse any act or course of action, so long as overall well-being is maximized through it. Samuel Scheffler (1988) conveys this concern as: “If overall satisfaction will be maximized under an arrangement in which goods and resources are channelled to people whose circumstances are already comfortable, while other people are allowed to languish in abject poverty, then that arrangement is precisely the one utilitarianism will recommend.” Translating this to enhancement: If overall satisfaction will be maximized under an arrangement in which welfarist enhancements are channelled to people who are already leading good lives, while other people are left leading unenhanced, horrible lives, then that arrangement is precisely the one utilitarianism will recommend. The result: profound unfairness and a potential violation of basic rights.

This possibility has been expressed in the context of enhancement by Savulescu and Birks (2001) through this hypothetical scenario:

One of the features of utilitarianism is that it aggregates well-being across individuals. This means that many small benefits may justify a great single harm. For example, imagine that genetic enhancement of intelligence becomes possible

and affordable. Assume that large numbers of normal people can modestly increase their intelligence and that this provides them with a greater opportunity to compete in a technology-rich society. A few cannot afford this and they become unemployable. Even if the harm to those who become unemployable is significant and the benefit to those who employ enhancement is small, utilitarianism would support enhancement if the numbers of the former are significantly smaller than the numbers of the latter.

This, it may be argued, would be an injustice. However, I believe the argument that utilitarianism would endorse any such scenarios is mistaken and rests on making what has been termed the wealthitarian fallacy (Greene 2014). To understand this fallacy, we need to examine the principle of diminishing marginal utility and how it entails that utilitarianism, contrary to this objection, has highly egalitarian implications. I do so in the next section, followed by an explanation of how the wealthitarian fallacy relates to it. With those concepts clarified, I will then return to this objection and attempt to show why it fails.

3. The principle of diminishing marginal utility

The principle of diminishing marginal utility (PDMU) is key to understanding how utilitarianism is egalitarian. As we will see, it entails that aiming to impartially maximize well-being will tend to also maximize equality. Being clear about what it is exactly that exhibits diminishing marginal utility will be central to showing how utilitarianism (and specifically sum-ranking) would not endorse basic rights violations and extreme inequality or unfairness. More relevant to my thesis, this will be key to showing how utilitarianism would favour a broadly egalitarian distribution to enhancements, one that generally prioritizes the worse-off, despite aiming to develop and distribute enhancements in whatever way that expectably maximizes overall well-being.

A good that exhibits diminishing marginal utility is one that confers less well-being to an individual the more that individual has of it (Greene & Baron 2001). Here, I use utility and well-being interchangeably. Income is a classic example: the utility of \$10 for someone with a million dollars will, generally speaking, be far lower than for someone with only \$1000. Similarly, for someone with one raggedy jacket in a cold environment, having one more jacket would, generally speaking, entail far more utility than an added jacket for someone with ten. To link this to functionalist enhancement, consider strength: someone who can barely hold themselves up before collapsing would generally benefit far more from gaining the strength to carry an extra 2kgs, than someone who has an average degree of strength and is able to walk and jump normally.

I emphasize “generally” in these examples because there will be exceptions. It is surely conceivable that a millionaire would derive more utility from \$10 than a poor individual who voluntarily leads a deeply ascetic existence. Similarly, it is conceivable that the individual with ten jackets has an obsession with jackets and on that basis they may derive more utility from it than the individual with one worn out jacket.⁴ Likewise, compared to someone with a debilitating leg condition but who has sedentary inclinations, it is possible to imagine an Olympic lifter gaining more utility than them through being able to lift or squat an extra 2kg.

Nevertheless, we can already see how PDMU entails that maximizing utility will tend to maximize equality as well. In general, the utility gained by someone impoverished if they are handed \$1000 will likely be far greater than the utility gained if that sum is given to a millionaire. More importantly, if we were to take those \$1000 from the millionaire and give them to the impoverished person, the gain in utility for the latter will likely far outweigh the loss of utility for the former. Similarly, the transfer of some political powers from an authoritarian, elite class to an oppressed populace with few rights will generally create far more utility than the loss of utility for the ruling class.

What these examples reflect is that, other things being equal, utilitarians should redistribute from the rich and powerful to the poor and powerless until the marginal well-being loss to the rich and powerful equals the marginal well-being gain to the poor and powerless (Lazari-Radek & Singer, 2014). However, utilitarians would also be sensitive to the side effects of any such redistribution process. For instance, taking money from the rich and giving it to the poor may have disincentive effects on the rich (by having their money taken from them) and the poor (by giving them money they have not earned). While this might temper just how radical a redistribution this would be, it is nevertheless unlikely to be anything like the status quo in most countries.

The same applies to our bodily and mental states. If we characterize a good life broadly, as one with some degree of pleasure, desire satisfaction, and a range of uncontroversial goods like autonomy, friendship and knowledge, then it is clear that, to the extent that we can distribute bodily and mental states through a healthcare system, we should distribute it in a way that maximizes the number of people capable of such a life. This would mean directing more resources towards those who are worse off in terms of bodily and mental states rather than those who are better off and already capable of such a life. I say more about this in section 4.2.

⁴ Although this clearly depends on how we characterize well-being. This example employs a simple desire-satisfaction account purely for illustrative purposes.

It is worth highlighting two assumptions behind PDMU. For PDMU to operate, we have to assume that comparing utility between individuals is meaningful. We may concede that this will not always be easy to do, but in principle we are assuming that it at least makes sense to say, for instance, “Alice will get more utility out of having shelter at night than a well-fed Bob will get out of a piece of chocolate.” Second, we are assuming that different people will derive similar levels of utility from a given level of material goods or political rights. That is, we assume that Alice would enjoy living the good life about as much as Bob would, and that their experiences of an impoverished life would be meaningfully comparable.

These are assumptions that, for some, might seem too big to justify the implications of a moral theory. What if no such meaningful comparisons are possible? It is an admittedly troubling thought, but note that this is a problem for any plausible morality. If we cannot meaningfully uphold uncontroversial claims like “for most people, a day of being tortured is worse than a day of being cuddled”, then the intuitions that help us evaluate any moral theory would be useless. If this is a problem, it is a problem for all plausible moral theories.

Nevertheless, utilitarianism’s legitimacy seems to particularly rely on such comparisons, especially being able to make them at least somewhat accurately. Consider this utilitarian claim: Given that an extra \$100 will make a big difference to someone who is poor, but not such a big difference to someone who is rich, we should, all else being equal, redistribute to the advantage of the poor person. Given the premises that PDMU rests on, the assumption here is that individuals have identical utility functions – that is, an identical ability to derive utility from a given good.

However, as Amartya Sen (1973) has argued, this may not always be the case. For instance, some bodily or mentally impaired individuals will be less efficient than non-impaired individuals at deriving well-being from resources: their utility functions will not be identical. And yet, if some individuals are better at transforming an extra dollar into well-being than others, utilitarianism suggests that we should give those individuals more money, all else being equal. Doing so is more efficient and therefore entails maximizing well-being.

I will respond to the issue Sen raises about impaired individuals in section 4.2, but for now, how should we respond to the problem that we can never be entirely sure that individuals have identical utility functions? Even if it is possible to work out the utility function of a given individual, it is unlikely to be possible to work it out for everyone. When we formulate a policy that will affect millions of individuals – say, a certain tax – we will not be able to take into account each individual’s utility function.

However, for policy purposes, we do not need to have knowledge of individual utility functions – we merely need to know that those functions conform to PDMU. Since we have no reliable way of comparing the well-being of individuals, it seems reasonable to base public policies on the assumption that all human beings – with the exception of, say, individuals in vegetative states (Singer, 2011) – have similar utility functions. That is, when it comes to big decisions that societies make, we do not need to know the well-being of a given individual or their capacity for it with much accuracy. Instead, we need to understand general patterns (Greene 2014): What kinds of policies tend to increase well-being? What kind of policies tend to decrease it? Such questions, based on the force PDMU, will tend to maximize equality.

This conclusion is crucial to my proposal of developing and distributing enhancements on the basis of utilitarianism. The obvious concern with developing and distributing enhancements that can expectably maximize overall well-being is that this may lead to extreme inequality or unfairness. If the best way to maximize overall well-being is to develop and distribute enhancements that favour a particular group, or that may neglect a certain minority, then utilitarianism would favour this. PDMU, however, shows such inequalities would not actually maximize overall well-being, although this is a point I return to below. We are now nearly set to consider objections against using utilitarianism to develop and distribute enhancements. However, it is worth highlighting that, despite how seemingly intuitive PDMU is, people (including philosophers) tend to have a difficulty thinking about it logically.

3.1 The wealthitarian fallacy

The wealthitarian fallacy is the description Joshua Greene (2014) gives for the tendency to conflate utility (or well-being) with wealth. Greene argues that many objections against utilitarianism rest on a mistaken understanding of utility that conflates the concept with “useful goods” – that is, the sorts of things that, if you have enough of, you may be considered wealthy or powerful (and in that sense, it includes liberties as well). As we have seen, useful goods (like wealth and political powers) tend to exhibit diminishing marginal utility: if you have very few useful goods in your life you stand to gain more utility from getting a little more useful goods compared to someone who already has lots of useful goods.

In contrast, utility itself does not exhibit diminishing marginal utility. Utility is a measure of the value we gain from having or consuming things. These are not limited to market goods, but can

include things like friendships, health and achievements and other contributors to, or constituents of, well-being. But it is impossible, by definition, to gain different amounts of utility from a given quantity of utility (Greene & Baron, 2001). It is worth rephrasing that in welfarist terms: it is incoherent to claim to derive different amounts of well-being from a given quantity of well-being. Yet this is different from claiming to derive different amounts of well-being from a given quantity of resources or goods.

Despite the seeming obviousness of this, many of us, including philosophers, tend to conflate the two when discussing scenarios that raise objections against utilitarianism. This is probably because we never have to make judgements about utility directly.

To illustrate this phenomenon empirically, Greene and Baron (2001) presented individuals with hypothetical societies that had varying distributions of annual income. For instance, in country A the bottom third's income might be \$25,000, the middle third \$45,000 and the top third \$70,000. In country B, the distribution of income is the same for the middle and top third, but the bottom third earns only \$15,000. If participants had an equal chance of living in either of these economic classes, which country would they prefer? Results showed clear preference for country A: the bottom third there has more income than the bottom third in Country B. Similar comparisons demonstrated that participants do not believe that such \$10,000 jumps are all equal: there is much more value in going from \$15,000 to \$25,000 than \$70,000 to \$80,000. That is, the participants certainly thought wealth exhibited PDMU.

In the second phase, participants were asked to assign a utility value for going from, say, \$15,000 to \$25,000, \$40,000 to \$50,000, \$70,000 to \$80,000, and so on. Those utility values were expressed using numbers between 0 and 100, with each increment having the same value. That means an improvement from 0 to 50 should be just as desirable as one from 50 to 100. To be clear: if the participants perceive going from \$20,000 to \$50,000 as having a utility rating of 0 to 50, then, given their demonstrated understanding of the diminishing utility of wealth, they will not rate the utility of going from \$50,000 to \$80,000 as going from 50 to 100. To go from 50 to 100, the utility gained would have to be as great as the utility gained from going from \$20,000 to \$50,000. Presumably, it would be closer to going from \$50,000 to making millions of dollars annually. As expected, the participants gave their rating with more weight to increases at the bottom end of the income scale.

However, and here we come to the third phase of the experiment, this logic did not carry over when participants were asked to rate the desirability of living in hypothetical countries described in solely terms of utility rather than income distributions. In this phase, participants were only

told the utility of each income bracket, not the income itself. In these scenarios, a jump from 0 to 50 in the utility of an income is the exact same jump from 50 to 100. Hence, if the participants were consistent, they should not place any more weight on jumping from incomes rated as 0 to 25 compared to ones rated as 25 to 50. However, they were inconsistent, treating these utility ratings as though they exhibited diminishing marginal utility, giving more weight to a jump from 0 to 25 compared to a jump from 25 to 50.

What this experiment shows is that, while people have no trouble distinguishing utility from wealth, and while they understand the relevance of PDMU, nevertheless, when asked to evaluate the distribution of utility, not wealth, they evaluate their desirability on the basis of PDMU. As Greene (2014) says:

In other words, people look at the jump from an income that gets a utility rating of 0 to one that gets a utility rating of 25 and they think, 'With so little utility to start with, that increase would make a very big difference.'

It is a sound conclusion if they were thinking about wealth, not utility. It is this conflation of the two that makes up the wealthitarian fallacy. Interestingly, the participants Green and Baron (2001) used included a handful of philosophy graduate students and professors. Now the question is whether those who believe utilitarianism could lead to extreme inequality and basic rights violations are committing it. I think they are, and in the next section I show how.

4. Objections to Utilitarianism

4.1 The distributive objection

Recall towards the end of section two I noted that utilitarianism does not provide any moral constraints on what can be done to an individual in the course of maximizing overall well-being. In the context of enhancement, limiting its distribution to an elite, or depriving a minority from it, or even enforcing it on people – all may be justifiable so long as it maximizes overall well-being. Let us examine this objection in more detail. While I gave the example from Scheffler (1988), and from Savulescu and Birks (2001) in the context of distributing enhancement technologies, this objection is perhaps most famously articulated by John Rawls (1971), who argues that utilitarianism fails to take seriously the distinctness of persons.

Rawls (1971) elaborates that as far as utilitarianism is concerned:

There is no reason in principle why the greater gains of some should not compensate for the lesser losses of others; or more importantly, why the violation of the liberty of a few might not be made right by the greater good shared by the many.

Hence, utilitarianism's failure to take seriously the distinctness of persons is reflected in neglecting those who are worse off so long as there is a net gain for other people – as opposed to the more acceptable trade-off where an individual suffers for some period in order to have a better life overall (such as working hard to enjoy their retirement, as opposed to working hard so someone else can retire early).

Similarly, and more relevant to the topic of enhancement and the example by Savulescu and Birks (2001), if new enhancements were not affordable by a minority, and this led to them being, say, unemployable and/or ostracized, utilitarianism would support this so long as the added well-being to the majority out-weighed the loss of well-being to the minority.

The central point of interest in all these scenarios is this: it seems it is possible to envision a scenario where an undesirable inequality (whether of wealth or liberty) exists, and where it can only be eliminated at the expense of aggregate wellbeing. In other words, we can envision cases where the only way to eliminate undesirable inequalities is through going against what utilitarianism recommends.

In what follows, I argue that this is false: utilitarianism would endorse such scenarios only if its goal was to maximize wealth not utility. Since it is concerned only with the latter, eliminating such inequalities will consistently increase overall wellbeing.

Consider Rawls's claim that utilitarianism provides no reason why the greater gains of some should not compensate for the lesser losses of others. Are those gains and losses in "useful goods" such as liberties and/or wealth or in well-being itself? I believe the force of this objection relies on thinking in terms of liberties or wealth. Let us consider wealth first: imagine two societies with equal total amounts of it, with one distributing that wealth equally amongst its population while the other does not. Utilitarianism, the implication goes, would give no particular reason to favor the more equal society. Hence, in choosing between an equal society and one where a despot maintains incredible wealth by leaving a population to languish in abject poverty – as Scheffler (1988) imagines it – utilitarianism would be indifferent so long as both have equal overall value.

But note the wealthitarian fallacy at work here. We are not talking about the distribution of well-being. If we were, then clearly utilitarianism would also favor the more equal society: stripping the despot of his or her wealth and sharing it with the oppressed population would increase

overall well-being: the loss of well-being to the despot will be far outweighed by the gain in well-being for the masses.

On the other hand, if we were imagining this society as having an unequal distribution of well-being rather than wealth, the scenario would be very different. Here, the scenario is that, despite a utilitarian redistribution of all the useful good, well-being itself is still poorly distributed. That is, in this case, goods have already been distributed in a way that takes into account equal consideration of interests and PDMU, and that aims to mitigate the negative impact on well-being that arises from class envy and inequality. This would have, as I argued, egalitarian implications in terms of how goods are distributed. Therefore, for this scenario to work, we must simply stipulate that, despite this egalitarian redistribution of goods, the society nevertheless has an unequal distribution in well-being. In such a case, it is true that utilitarianism would be indifferent between such a society and another society with the same amount but more equal distribution of well-being. Both would be considered equally good. However, note how difficult it is to conceive of such a society where well-being is maximized *and* unequally distributed: sure, we may simply stipulate that this happens to be the case, but this now sounds less convincing. What exactly is this inequality in well-being arising from, given the egalitarian distributions that utilitarianism has already put in place? If we cannot identify an oppressive or unfair force in the society (such as a despotic regime or a market that is primed to favor an elite), then it is not clear that this inequality in well-being is relevant to justice or in any way objectionable. More to the point, it is difficult to plausibly envision such a scenario.

Now let us consider liberty: Rawls argues utilitarianism could allow for the violation of the liberty of a few so long as that is made right by the greater good shared by the many. True, utilitarianism would endorse curtailing some individual's liberties for the greater good shared by the many – for instance, stopping dangerous individuals from roaming freely or punishing drunk drivers. Any other plausible moral theory would curtail liberties in similar ways. Presumably, what Rawls has in mind is a greater violation in liberties. What might a relevant example be? Rawls (1971) considers an “extreme case” of a slaveholder justifying the status quo to his/her slaves on the basis of utility: it just so happens that in their societal circumstances, slavery is necessary to maximize well-being.

Now, it is true that, if this were the case, utilitarianism would support it – and for Rawls and other objectors to utilitarianism, they are free to stipulate whatever conditions they want for their hypothetical scenarios. However, it seems the more relevant question when met with these “extreme cases” is whether what they stipulate is plausible: could slavery, or something

like it, ever actually maximize overall well-being? For this to happen, each slave owner must, on average, gain more well-being from owning a slave than a slave loses by being one. This is not plausible. Greene and Baron (2001) illustrate this well:

Imagine a society in which half of its members are enslaved. That would mean that you, as a prospective member of this society, would have an equal chance of being a slave or a slaveholder. Now ask yourself (as a non-slave, non-slaveholder), which change would be more dramatic, the increase in happiness you would experience from acquiring a slave (assuming you have no moral problems with doing so) or the decrease in happiness you would experience by becoming one? There's no question. Slavery is much worse for the losers than it is better for the winners.

We will reach the same conclusion when considering a slave-based society using whatever slave to slave owner ratio we like. For instance, consider a society with a ratio of nine slaves to one slave owner. Which would yield a greater overall well-being: not being either a slave or slave owner, or have a 90% chance of being a slave and 10% of owning one slave? Clearly not being either a slave or slave owner maximizes overall well-being. Alternatively, let's imagine a ratio of one slave to nine slave owners. Which would yield a higher expected utility: not being a slave or a slave owner, or having a 10% chance of being a slave and a 90% chance of being a slave owner but sharing one slave with eight other slave owners?

No matter the ratio, it is impossible to see how slavery could ever maximize overall well-being.⁵ Of course, slavery or highly oppressive laws *could* maximize wealth or useful good more generally. If the goal of utilitarianism was to maximize wealth rather than well-being, then it is certainly conceivable that the loss of wealth and liberty from becoming a slave or poor worker would be outweighed by the increase in wealth from becoming a slave owner or tycoon. By exploiting slaves, slave owners could increase their wealth so drastically that it would outweigh the loss of wealth to becoming a slave. In that sense, utilitarianism could endorse the greater gains in wealth (of slave holders) for the comparatively lesser losses of wealth to slaves. But utilitarianism does not count wealth, it counts well-being. And while PDMU applies to wealth, it does not apply to utility. To conflate the two is, I have argued, to commit the wealthitarian fallacy.

And yet this conflation is recurring. While I have described it in terms of wealth and liberties, let us see how it may relate to enhancement. Recall the example used by Savulescu and Birks (2001) where the genetic enhancement of intelligence is possible and large numbers of normal people

⁵ Of course, this is presuming that to be a slave *is* to lead a bad, if not horrible, life. If we alter these scenarios to ones where being a slave is not actually a disaster to one's well-being, then perhaps we would reach different conclusions. Here, I use "slave" to mean a deeply oppressed person – which I presume is someone who is not capable of leading a good life, or at least has very low chances of doing so.

modestly increase their intelligence to have greater opportunities to compete in a technology-rich society. For the few who cannot afford this enhancement, they become unemployable. Utilitarianism, Savulescu and Birks (2001) assert, would support this even if the harm to those who become unemployable is significant and the benefit to those who use the enhancement is small – so long as the numbers of the unenhanced are significantly smaller than the numbers of the enhanced.

The first thing to note here is that it is not entirely clear what “benefits” and “harms” refer to. Do they mean the benefits of being more competitive, and therefore having access to better jobs/incomes and other material goods? Conversely, do they mean a loss of such access when referring to the harm that the unemployable undergo? Or are they referring directly to improvements and diminishments to well-being itself? As I noted in the examples above, there is a crucial difference. Let us presume they are talking about benefits and harms of having access to greater or lesser goods (which is implied by their reference to benefits as “greater opportunities to compete”).

To help us imagine what a “significant harm” to the unemployable might be, let us say that they went from leading a life with sufficient access to a range of goods, to a deeply insecure one where they struggle to even attain food and shelter. To help us imagine what a “small benefit” might be to the majority who have been enhanced, we can imagine that they went from leading a similarly decent standard of living, to one where their disposable income increases by, say, 10% (or whatever other access to goods such a raise might entail). Here, Savulescu and Birks (2001) seem to be saying that utilitarianism would support such a scenario if the numbers of the severely harmed are significantly smaller than the number of those marginally benefitted.

But note how this ignores PDMU. If those who are marginally benefitted are so much more numerous than the significantly harmed, then taking a tiny fraction of each of their new incomes and giving it to the harmed would mitigate that harm and create more utility overall than the loss of utility to the enhanced. To give a concrete example, imagine 90% of a population become enhanced, gaining a 10% increase to their incomes. In turn 10% become unemployable and lose most or all of their income. A distributive system based on utilitarianism would not allow for this to happen to begin with. It will always maximize utility to reduce some of the income of the majority and give it to the minority (for instance, through taxes) in order to more significantly increase overall well-being.

Now suppose what Savulescu and Birks (2001) had in mind was well-being, not wealth, when discussing the harms and benefits to the individuals in that society. In such a case, this would

mean that PDMU as well as the adverse personal and economic effects of envy and inequality have already been taken into account. Whatever redistribution PDMU would entail has already happened. In terms of access to goods, the minority whose well-being has been significantly diminished has already been compensated on the basis of PDMU – that is, in terms of material goods, there is no undesirable inequality to point to as the reason for their diminished well-being. The diminished well-being is simply there for reasons unrelated to wealth distribution. What might still be causing the diminished well-being now appears unobvious – if an example could convey such a hypothetical scenario, I suspect few would find it unjust. Such a scenario would entail that the unequal distribution of well-being was due to blameless forces: perhaps some individuals are merely predisposed to lower levels of well-being in a way that society cannot do anything about it through distributing resources in a different way.

Again, I think Savulescu and Birks (2001) are conflating a loss of access to goods (for which PDMU applies) with a loss of well-being (which is what we get after we apply PDMU if we are thinking in utilitarian terms).

Despite these clarifications, critics of utilitarianism might respond with ambivalence towards this heavy reliance that the theory places on PDMU to remain plausible. Utilitarianism provides no principled guarantee that it will look out for the vital interests of individuals – all it can say is, given how we *happen* to derive utility from resources, we do not need to worry about utilitarianism endorsing the curbing of vital interests of some for the greater good. This seems not good enough. For Rawls, utilitarianism's failure to take seriously the distinctness of persons is ultimately about its inability to provide a *guarantee* that it will not violate basic rights. Utilitarianism can only appeal to PDMU to show that it's implausible it would ever sacrifice the vital interests of some for the greater good – but that is not a guarantee. Rawls's theory, with its emphasis on the least advantaged, provides such a guarantee – so why, standing behind the veil of ignorance, would you ever accept this risk of utilitarianism?

It is worth emphasizing two points here. First, no plausible ethical theory actually provides a guarantee that it will not violate the vital interests of some for the greater good. Only the most absolutist deontologist would think it is wrong to sacrifice one human life in order to save all other humans – an approach to ethics that is challenged on several fronts (Waldron, 2011). Yet, as soon as we concede that it would be right to kill one human (or otherwise deprive them of their vital interests) in order to save all of humanity, then we have also conceded that, at least in some highly unlikely scenarios, no such guarantees hold. Once we accept this, it should be clear that when we speak of guarantees for the protection of individuals' vital interests, we don't

really mean a guarantee – we mean a guarantee for most cases, we mean a guarantee in the real world where we do not need to consider such unlikely scenarios.

In that way, utilitarianism does provide a guarantee – at least one that is as good as we can hope to secure. PDMU is not a superficial basis upon which to make a theory plausible. If humans were not the sorts of creatures that derived utility from goods in a way that generally accords with PDMU, then we would be fundamentally different creatures. Every notion we have about well-being and fairness is linked to this principle. If that principle were different – if we were the sorts of creatures that, say, derived more and more utility with every extra dollar or spoonful of ice cream – then we can be sure that whatever intuitions we have about right and wrong would also be unrecognizably different. Indeed, it is not entirely clear how to even conceive of such fundamentally different beings.

Ultimately, this reliance on PDMU is not to suggest that the imagined scenarios used against utilitarianism are so unlikely to occur that we can afford to ignore them. Instead, the goal here has been to be clearer on what utilitarianism would actually endorse and to correctly imagine the hypothetical scenarios its critics evoke to mind. Doing so should make it clear that, in general, aiming to maximize well-being tends to maximize equality as well.

4.2 Do human capacities exhibit diminishing marginal utility?

In the examples above, I applied PDMU in the context of goods like wealth and liberties; even in the Savulescu and Birks (2001) example, while it invoked genetic enhancement, what was at stake was the effect of those enhancements on people's wealth and how material goods were distributed in that society more generally. However, since our concern is about the distribution of enhancements – that is, the advantageous bodily and mental changes themselves, or the interventions associated with them – it seems relevant to ask whether bodily and mental states are the sorts of “goods” that exhibit PDMU. Do things like sensory acuity, intelligence and immunity share the same relationship with utility as, say, income? The one example I used above relating to a gain in strength equivalent to being able to carry an extra 2kg suggested that strength in general displays diminishing marginal utility. It seems plausible that, in general, someone with muscular dystrophy who is cured of the condition will derive far more utility from that bodily change than someone of average athleticism suddenly having elite athletic ability.

Indeed, it is possible to imagine that at some point, having ever greater athleticism would confer disutility in the form of being too strong for one's own good.

The same would hold for more complex capacities such as intelligence. Having an IQ below 70 makes it extremely difficult to attain many of the things associated with a good life: being autonomous, having meaningful relationships, achieving things of value, being knowledgeable, and so on. On the other hand, having an IQ over 120, all else being equal, is arguably more than enough to be able to attain these goods. Hence, if we think of intelligence as being instrumental to having access to these constituents of well-being, then it is clear that it exhibits diminishing marginal utility: going from an IQ of 70 to 80 would entail a far larger jump in one's ability to attain these goods. But going from 120 to 130 (in and of itself) is unlikely to do much in that regard, and indeed, it seems plausible that at some point purely augmenting someone's intelligence, or some component of it, will backfire. For instance, cases of hyperthymesia, where an individual has extremely detailed autobiographical memories, are marked with significant life disruptions (Parker, Cahill, & McGaugh, 2006).

What this suggests is that, in general, utilitarianism would promote the distribution of enhancements in a way that prioritizes the worse off: all else being equal, improvements in their capacities are likely to generate more utility than improvements in people with average or above average capacities⁶. In fact, utilitarianism would promote the distribution of enhancements in a way that is far more intuitive than some of its critics would recommend.

It is worth giving a brief example of this using an example of genetic enhancements. In a hypothetical society where we can select for genes, Farrelly (2002) argues for a Rawlsian distributive system. While Rawls was concerned with the just distribution of "social primary goods" (rights, liberties, income, and self-respect), the prospect of genetic enhancement may make it possible to also determine people's "natural primary goods" (health, intelligence, vigour, imagination). In a Rawlsian spirit, Farrelly suggests that genetic enhancements should be distributed using a "genetic difference principle" where:

Inequalities in the distribution of genes important to the [natural primary goods] are to be arranged so that they are to the greatest benefit of the least advantaged.

That is, we should distribute genetic enhancements in a way that would maximize the genetic profile of the least endowed. Let us imagine these genetic enhancements are welfarist: they will

⁶ As I noted in the cases of George and Lennie, there may be exceptions. Utilitarianism needs to look not only on the well-being impact of a bodily or mental change for that person, but also how much well-being that change will enable them to generate through what they do with that change.

actually increase the well-being of people who undergo them. Based on this, consider a hypothetical society's present genetic distribution:

Genetically...	% of population	Individual well-being level
Best off	10%	80
Middle	80%	20
Worst off	10%	19

Here, those who are best off genetically have far more well-being than those in the middle and the worst off – indeed, there is little difference between those in the middle and the worst off. A suite of genetic enhancements, unaffordable by the worst off, can change how much well-being the people of this society have in the following way:

Genetically...	% of population	Individual well-being level
Best off	10%	85
Middle	80%	80
Worst off	10%	19

Here, the worst-off do not undergo these enhancements and do not benefit from them. However, the middle benefit hugely and the best off marginally so. Should we support such a change?

For utilitarianism, this is a no-brainer: clearly well-being has improved overall through the introduction of the enhancements, from a total of 119 to 184. This would apply regardless of whether we are talking about genetic enhancements or some other type of enhancement. Yet for the Rawlsian, we should not move towards such a distribution, as it does not benefit the worst off. However, note that we are talking about well-being here, not wealth, power or any other good that exhibits diminishing marginal utility. That is, for utilitarianism, these are the figures we get *after* we have already taken into account PDMU as well as envy and whatever resulting class tension such a distribution might engender. That is, this is not a description of a society where the worst off are materially or politically marginalized to the benefit of the rest. Nevertheless, the Rawlsian would not favour this latter distribution, despite how it greatly

improves the well-being of 80% of the society (who were barely doing better than the worst off), while also modestly improving the utility of the best off – all without leaving anyone worse off. Clearly the utilitarian provides a much more plausible response than the Rawlsian does here.

Ultimately, in arguing that our capacities exhibit diminishing marginal utility, it is worth making a distinction between interventions in the body or mind that enable a person to derive more utility by gaining access to more goods, and interventions that augment their capacity to derive utility from those same goods. The distinction can be seen in the difference between an intervention that enables one to get a job, and therefore money, and therefore other goods, and an intervention that enables one to derive more utility from those same goods. Examples might be a deeper sense of gratitude, a sharper intellect or senses. Individuals who have undergone such changes may have weightier interests than those who have not. Their sharper intellect and senses may mean they have a greater interest in avoiding pain than others, for instance, and they may gain more utility from the same meal, book, or friendship than someone whose intellect and senses have not been similarly sharpened. That is, they have undergone changes to their biology and psychology that increase their chances of leading good lives by improving their very ability to derive utility from goods. If we frame this in economic terms, such enhanced individuals may derive more utility from \$1 than the average unenhanced person. If so, utilitarianism seems to favour devoting more resources to them: they are better at transforming those resources into utility and hence, in order to maximize well-being, prioritizing these enhanced individuals may be the utilitarian thing to do.

This, I believe, is a misleading picture of utilitarianism. The difficulty arises from trying to imagine what such utility-deriving enhanced individuals are like. However, as noted in passing in section three, Sen makes an argument against utilitarianism in terms of how it would treat people with physical or mental impairments. If impaired humans are not as efficient at converting resources into well-being, utilitarianism would direct resource away from them and to people who are more efficient. In the scenario with such enhanced people, the unenhanced may become the equivalent to impaired (Wolbring 2006) and hence marginalized. This seems troubling for utilitarianism and is worth investigating in more detail: how would utilitarianism actually treat impaired individuals?

4.3 Utilitarianism and people with impairments

Some have argued that utilitarianism might neglect individuals with impairments that reduce their capacity to derive utility from goods compared to others. For instance, Sen (1973) argues that, in aiming to maximize well-being, utilitarianism would favour giving more goods like income to individuals who are better at deriving utility from it. He calls this ability an individual's welfare function, and says this would be an unacceptably unjust conclusion reached by utilitarianism. Is Sen right?

Let us explore in more detail what Sen has in mind. Sen (1973) critiques utilitarianism's "ill-deserved egalitarian reputation" by asking us to consider "a case where one person A derives exactly twice as much utility as person B from any given level of income, say, because B has some handicap, e.g., being a cripple." This means that A is twice as efficient at converting income into utility compared to B. Given that utilitarianism would aim to maximize the sum-total of utility of A and B, it would require that person A be given a higher income than B. This, the argument goes, shows utilitarianism's true character: Not only does the theory ignore the inequality between A and B, it compounds it by giving more income to A, who is already better off. Thus, Sen (1973) concludes:

The utilitarian maximand [i.e. it's maximizing goal] discriminates against a person who is uniformly handicapped in converting income into utility (since she would be seen as an 'inefficient' utility maker, with a low utility-generating ability). The utilitarian logic is insensitive to the fact that giving her less income would compound the lowness of her utility-generating capacity: she would get a lower total income in addition to having lower utility per unit of income.

This, at first sight, seems a compelling argument, and in the context of enhancement would translate as the unenhanced person taking the place of the "uniformly handicapped" here. However, a recurring observation in this thesis is that abstract examples often lose their plausibility once we make them more concrete. So let us consider how Sen's person A and B might translate into the real world. What would it take for an impaired person to be half as efficient at converting income into utility as a non-impaired person? I will argue that, for this to be the case, the impaired person would have to suffer from a condition entirely different than the one Sen imagines.

Nevertheless, let us consider the case of a paraplegic. How could being paraplegic render someone less efficient at converting income into utility? Let us imagine their income is equal to someone who is able to walk. In terms of the utility derived from being able to afford food and

shelter, there is no reason to believe they are any different. The same applies to their appreciation of film, music, books and travel. It is true that the non-disabled person might have a preference for those and therefore derive more utility out of these, but it is equally possible that the paraplegic has the same, if not more, capacity. The fact that someone is impaired does not necessarily mean that their (overall) ability to derive utility from resources is compromised.

But note how much more utility the paraplegic person can derive from, say, being able to afford a motorized wheelchair⁷. The utility derived from being able to afford one is arguably greater than any amount of utility that a non-disabled person can derive from the use of their disposable income.⁸ To the extent that this is true (something that can only be determined by the specifics of the example), we can easily argue that, in fact, the paraplegic is *more* efficient at converting income into utility. Depending on the resources available in their society, enabling the paraplegic to afford a motorized wheelchair should be prioritized over allowing the non-impaired person to afford comparatively non-vital goods as the good of mobility. More broadly, given how much resources that ameliorate impairments can greatly benefit the disabled, utilitarianism would endorse the redistribution of resources from the non-impaired to the impaired in order to provide those benefits (Stein, 2002).

To be fair, this example does not play by the rules set out in Sen's example: the paraplegic is *not* in fact half as efficient as the non-impaired person at converting income into utility (in fact, they are potentially more efficient). I justify breaking the rules of Sen's scenario only to illustrate the implausibility of someone with an impairment such as this one being so inefficient at income-utility conversion.

It is difficult to imagine what exactly it might mean to derive half as much utility from resources as a non-impaired person, and how that might relate to income. But consider this example: One person, call them Leibniz, has a normal capacity for tasting and appreciating fine food, while his colleague Newton has a significantly diminished capacity for taste. There are only two visually indistinguishable dishes left in the dining hall they both just entered for supper. Both dishes are of equal nutrient and caloric value, except one is quite bland tasting while the other is imbued with subtle, delicious flavours. Here, even though Newton's impairment leaves him worse off (we can presume), it nonetheless would not be unfair to give the finer tasting dish to Leibniz,

⁷ I borrow this example from Stein (2002)

⁸ To the extent that we may be worried about our ability to make such interpersonal utility comparisons, it is often helpful to turn the scenario onto ourselves. In this case, you might ask what would be more valuable to me: would having some extra money to spend on non-vital goods make up for the fact that I can't easily travel to other places? The answer, for most people, will be clearly "no".

who can actually appreciate it. Here, utilitarianism conforms with what I suspect is most people's intuitions about fairness: Newton is not able to appreciate the flavoursome dish, so why give it to him?

Similarly, we can think of an example where only two rooms are available, one with a spectacular view, and another facing a brick wall. If Newton's sight was poor and entailed that he would only perceive a blurry mess from either window, while Leibnitz would actually be able to appreciate the splendid view, it is not clear why it would be unfair to give the flat with the better view to Leibnitz (all else being equal).

What these examples suggest is that, if someone's ability to derive utility from certain resources is deficient, why would it be unfair to direct those resources to individuals who are better able to? But here, we are talking about specific resources. When it comes to income, a generic resource for attaining all sorts of other resources, a certain diminished capacity will not be enough to justify giving the impaired person a lower comparative income. Newton, despite limited taste or vision, can still enjoy a great many other things just as well as Leibnitz, if not more.

Nevertheless, these are examples where someone who is functionally better off is favoured over someone who is worse off. And yet, while this may not be egalitarian in the sense of giving people equal *treatment*, it is egalitarian in a more significant way. To illustrate how, Peter Singer (2011) gives this example:

Take a relatively straightforward example of an interest, the interest in relief of physical pain. Imagine that after an earthquake I come across two victims: one with a crushed leg, in agony, and one with a gashed thigh, in slight pain. I have only two shots of morphine left. Equal treatment would suggest that I give one to each injured person, but one shot would not do much to relieve the pain of the person with the crushed leg. She would still be in much more pain than the other victim, and even after I have given her one shot, giving her the second shot would achieve a more marked reduction in her pain than giving one shot to the person in slight pain would do for that person. Hence, equal consideration of interests in this situation leads to what some may consider an unequalitarian treatment: two shots of morphine for one person and none for the other.

Here, an unequal treatment of two individuals not only leads to greater overall well-being, but it leads to a more equal result: by giving the double dose to the more injured person, we bring about a situation where both individuals are experiencing more similar levels of pain. Instead of ending up with one person in considerable pain and another in no pain, we have ended up with

two people in slight pain. Here, utilitarianism clearly favours what I suspect most would find a fairer outcome.⁹

Linking this back to our enhanced utility-deriving individuals with their sharper senses and intellect and where the unenhanced are equivalent to the impaired here, the above discussion would lead to the same conclusion. That is, the unenhanced, like the impaired, do not necessarily derive any less utility than the enhanced overall, and to the extent that they do, they may derive more utility by directing resources towards them that enable them to sidestep their unenhanced limitations. On the other hand, to the extent that we agree that it would not be unfair to give Leibnitz the more delicious meal and the room with the view, it is not clear why it would be unfair to favour the enhanced person in similar ways, all else being equal.

This qualifier of “all else being equal” is important: if the enhanced person is enhanced because of some unjust reason, or if the unenhanced person wants to be enhanced but cannot because of some unjust reason, or if this policy more generally led to resentment and social strife of some kind, utilitarianism would take this into account to the extent that they have implications for overall well-being, and hence perhaps reach different conclusions.

One final issue to consider here is individuals with impairments that truly diminish all their capacities to derive utility from anything – say, individuals in vegetative states or with severe intellectual impairments. Utilitarianism suggests that it is permissible (and may indeed be required) to direct less resources to them than to others who can actually utilize those resources. Again, it is not clear what is objectionable about such a conclusion, although clearly much will depend on the details of a given scenario.

To sum up what I have concluded so far in this chapter, I have argued that utilitarianism would favour promoting the development and distribution of welfarist enhancements that can expectably maximize overall well-being, which I have described as the combined prudential and moral value of an intervention. When it comes to developing or distributing an enhancer, two questions are key: Will its development expectably maximize overall well-being? And if the enhancement already exists: In what way should it be distributed in order to expectably maximize overall well-being? I suggested that it is difficult to see how the well-being derived from enhancers that augment capacities can compete with vital treatments, but that this picture is complicated by the possibility that the former could enable individuals to do far more for

⁹ Of course, other distributive theories might agree too – and to the extent they do, this would be an instance of utilitarianism reaching broadly uncontroversial conclusions about what to do in real-world scenarios.

overall well-being (such as through inventing, say, life-saving technologies). A utilitarian society will have to weigh up these possibilities and distribute enhancements accordingly – this is something that I then argued will not lead to severe inequalities or violations of basic rights, nor marginalize the unenhanced or impaired.

So far, this discussion has been formulated with the purpose of examining how practical ethicists and policy makers might use utilitarianism to think about the development and distribution of welfarist enhancements. However, in reaching these conclusions, a crucial question arises: how do they factor in the real world, where organised political communities (states) are broadly responsible for how we distribute goods? In the next section, I consider this issue.

5. Utilitarianism, the state, and enhancement

The state plays a large role in determining how goods are distributed and what goods are developed – even in the context of free markets, the state formulates laws on what can be legally produced and consumed, and regularly interferes with the market to help ensure certain outcomes (Chang, 2011). Hence, whether enhancements are developed or distributed publicly or privately, the state plays a pivotal role in that process. Because of that, and to make this discussion less abstract, it is worth investigating how a state might use utilitarianism to distribute enhancements. Any full examination of such a prospect would easily be a thesis of its own. Here, I only provide a brief investigation of this topic. My goal is merely to convey the point that the state using utilitarianism to develop and distribute enhancements, while worth exploring, is naïve to the real world. That is, trying to apply a normative theory to the workings of the state can be a useful exercise in ideal theory, but it is unlikely to translate into anything actionable in the real world. I show why at the end of this section before proceeding to the final chapter to explore a more practical use for the utilitarian development and distribution of enhancements.

So how should a state distribute enhancements from a utilitarian perspective? The ideal answer is: in whichever way that maximizes overall well-being. However, well-being is not something that can be directly distributed. Firstly, for something to be distributable, it must be tangible and, in a sense, be capable of being detached from people, moved between them, and allocated to them – like money. Secondly, for something to be quantifiable, it must be epistemically accessible to second parties: others must be able to see how much a person has of it – again, like money. To the extent that any plausible account of well-being has a subjective component

to it, well-being fails on both these accounts. It is not a good that can be directly allocated and quantified. Given these constraints, a state might try to only distribute the means to well-being.

What are those means likely to be? As I've previously suggested, an overlapping consensus about the basic contributors to well-being are likely to involve various features of each theory, with a list that includes items such as pleasure, health, autonomy, some degree of desire satisfaction, and friendships. Again, this is not a list that claims to be objective but merely one that holds these as reliable contributors that are likely to be instrumentally or non-instrumentally valued by a plausible theory of well-being.

However, even these means to well-being suffer from the same problem well-being does: they are not the sorts of things that can be distributed or assessed by second parties to determine to what degree they exist in a person's life. Even health has a subjective component (at least in regards to mental health). This entails that something other than the means to well-being needs to be distributed: these are proxy goods or opportunities that we have reason to believe are reliable contributors to the means to well-being.

What are those proxies? There are two ways to think about these. One is that they are resources: goods that are known to reliably contribute to the well-being of the person who has access to them. Examples are healthcare (not health), education (not knowledge), income, leisure time, housing, and so on. We can think of an approach to distributing these as "resourcist". Another way to think of those proxies is in the form of capabilities. These refer to individual people's freedom and ability to achieve well-being through real opportunities to do and be what they want (Robeyns, 2011). Examples include health, senses, imagination, emotions, practical reason, and control over one's environment (Nussbaum, 2011).

Measuring the capability sets of a person is likely to offer a more accurate proxy of their actual well-being. However, to the extent a capabilities approach may track well-being more accurately, it arguably also faces the same problem of trying to measure the means to well-being: for policy-purposes, how do we measure the capabilities of individuals, and how do we generalize from what a sample in a population has in terms of capabilities to the rest of the population? A capabilities metric seems far harder to operationalize for policy-purposes (Pogge, 2010; Robeyns, 2006)

Because of this, I favour using a resourcist approach, where certain standard goods like the ones identified above act as proxies for the likely amount of well-being a person in a society has. These standard goods may differ from state to state, or at least how they are weighed in relation

to each other is likely to differ. Here, utilitarianism would instruct each state to come up with a weighted list of standard goods where that list's general acceptance by a population can be expected to maximize its overall well-being over the long-term.

It is worth highlighting at this point the compatibility utilitarianism shares with sufficientarianism, not as an ethical theory, but as a pragmatic approach to identifying the sorts of goods that should be publicly funded. This is because utilitarians can argue that any society based on the principle of utility ought to provide a social welfare net that provides everyone with the sufficient means to access goods like adequate food, shelter, education, and healthcare. This is because few things cause more avoidable suffering than being unable to meet such basic needs (Lazari-Radek & Singer, 2014).

Hence, imagine a state uncovers that a list of resources like the one below, ordered on the basis of priority, is expected to maximize citizen's overall well-being in the long-term:

- Food
- Healthcare
- Clothing
- Housing
- Safety
- Education
- Income
- Clean environment
- Leisure time

Based on this, an effective way to maximize well-being is to publicly fund these goods, at least for certain demographics, so that individuals who cannot afford them can have access to them to a sufficient degree. Here, what a sufficient degree is need not be arbitrary: utilitarianism is employing a threshold for public funding not on the basis that sufficiency is the correct approach to ethics, but as a pragmatic approach to minimizing avoidable suffering over the long-term. What that threshold for sufficiency is will be determined in the same way the list is: given a state's resources and its citizens' particular needs, a threshold would be set based on what can be expected to maximize overall well-being in the long-term.

Of course, the picture gets more complicated by the fact that each of these items will have its own list of constituting items. For instance, healthcare would include infrastructural issues like the number of hospital and clinics, as well as interventions, with interventions themselves being

prioritized in terms of public funding or their development based on their likely contribution to overall well-being. For most states, this would mean publicly funding at least some enhancements, especially ones that are the most reliable proxies to well-being, such as life-saving medications and palliative care. As for other enhancements, if a state has the resources, it may cover enhancements that are likely to be less significant contributors to well-being, such as transmedical enhancements. In contrast, in a poor state with few resources, this may be difficult to justify.

For instance, according to the World Bank (2015) France's GDP per capita is around US\$36,000. It spends nearly 12% of its total GDP on healthcare to ensure that all its legal residents are largely reimbursed for most of their medical needs. Life expectancy is 82 years. Within such a context, it may be that the benefits of publicly funding transmedical cognitive enhancements would offset any costs associated with paying closer to 13% or 14% of its total GDP in order to publicly fund them. Contrast this with Nigeria's US\$2600 GDP per capita (World Bank, 2015). Nigeria spends around 3.7% of its total GDP on healthcare where life expectancy is 53 years. It has a number of endemic diseases, a high prevalence of HIV, and has some of the highest maternal mortality ratios in the world, with over 70 deaths per 1000 live births, compared to France's 4 (World Health Organization, 2015). Within such a context, it is unlikely that publicly funding, say, effective pharmaceutical cognitive enhancements would maximize well-being most efficiently. Having said that, it may be that France, rather than funding such cognitive enhancers for its citizens, should use these resources to help improve well-being in Nigeria instead.

This last suggestion makes clear the problem with trying to impose an ethical theory onto the workings of states: any such proposal will be limited by its neglect of real world politics, and the fact that countries with democratic and stable institutions are a minority (Zohny, 2014b). States are not necessarily committed to maximizing their citizen's long-term overall well-being, and they certainly are not committed to impartially maximizing overall well-being across the world. Even in a nominally liberal democracy like the US, evidence is overwhelming that economic elites and organized groups representing business interests have substantially more impact on government policy compared to average citizens (Gilens & Page, 2014). While what qualifies as a democracy is contestable, according to the Democracy Index (The Economist Intelligence Unit, 2015), out of 167 states in the world, 56 are "flawed democracies" (and that does not include the US, which is considered a 'full democracy'), while 37 are considered a hybrid between democracy and autocracy, and 51 are run by authoritarian regimes. Hence, even if we presume that democracies are able to reflect the will of their citizens and that that will is one of institutional utilitarianism, it is clear that most people do not live in such states.

Of course, utilitarianism can still be made to work within these constraints, accepting this political reality while trying to maximizing overall well-being within it. But it is clear that, given the preceding paragraph, such an endeavour is unlikely to see anything like enhancements being developed or distributed in ways that expectably maximize overall well-being across the world, as opposed to the state. Because of that, I propose a different approach that sidesteps the limitations of the state, which is the topic for the next and final chapter.

6. Conclusion

I argued that utilitarianism is particularly well-suited as a normative guide to public policy, appealing to well-being as a common currency that can help resolve conflicts between other values. Its three components have an intuitive appeal that sets them up as good default position for a metamorality. Understanding how those components lead to equal consideration of interests, along with what PDMU entails and the distinction between wealth and utility, we can see how many objections against utilitarianism are mistaken: they illogically apply PDMU to utility or misconstrue how utilitarianism would respond to unenhanced or impaired individuals in concrete examples. When it comes to enhancements, utilitarianism would prioritize those on the basis of how vital they are to overall well-being. In the context of a given state, utilitarianism would look for reliable proxies to well-being and would place enhancements depending on the priority healthcare is given, with enhancements that are more reliable contributors to well-being being more likely to be publicly funded.

Having reached these conclusions, the looming problem for this discussion is that it has unfolded in the realm of ideal theory – the real world, along with the fact that it is unlikely societies will start adopting utilitarianism, raises doubts about how actionable any of the conclusions I reached here are. This is why, in the last remaining chapter, I propose a mechanism that relies on incentivizing innovators to develop and distribute enhancements that expectably maximize overall well-being across the globe, while also mitigating the possibility that enhancements may exacerbate existing inequalities within and between societies. While this mechanism is founded on utilitarian principles, I will argue it can also appeal to non-utilitarians as well.

Chapter 7

Enhancement & Inequality: The Well-Being Impact Fund

This chapter brings together the different strands of this thesis into a concrete proposal. It is a proposal that deploys the welfarist account of enhancement; that is rooted in a utilitarian approach to the development and distribution of enhancement; and that, ultimately, works to mitigate the possibility – or, I will argue, the likelihood – that enhancement technologies may exacerbate existing inequalities and/or lead to neglecting vital interventions that combat disease.

While I alluded to the possibility of enhancements exacerbating existing inequalities in the introduction to this thesis, I have not provided an argument for it so far. I do so below. First let me link this to some of the conclusions reached in previous chapters. These are as follows: As emphasized towards the end of chapter five, the welfarist account compels us to assess and respond to the possibility that some interventions (namely, putative functionalist enhancers) may reduce our chances of leading good lives – it compels us to do so because, in order for such an intervention to *be* enhancing on the welfarist account, it must in fact tend to increase our chances of leading good lives. Hence, the account directs us to engage with the possibility that, for example, some interventions may reduce our autonomy or authenticity or capacity to lead meaningful lives, potentially reducing prudential value in our lives. Secondly, while the account responds to concerns of this prudential kind for users of an intervention, it does not concern itself with other ethical implications of enhancements related to, say, their fair distribution. For this I invoked utilitarianism, which, despite allegations that it may in fact completely fail to respond to this concern, I argued will tend toward an egalitarian distribution of enhancement, while also acting as a metamorality that adjudicates between different moral stances. However, the remaining question at the end of the previous chapter was how do we put this into practice in the real world, with all its political obstacles and the imperfections of the state. This brings us back to my proposal, which I call the Well-Being Impact Fund.

The Well-Being Impact Fund builds on an existing proposal – the Health Impact Fund (Hollis & Pogge, 2008) – which aims to counter some of the drawbacks arising from the use of patents in pharmaceutical innovation. Specifically, amongst its aims is to reduce global disease burden by preventing high prices for essential drug treatments and ending the neglect of diseases of poverty. My proposal supplements these aims by working to prevent high prices for valuable

transmedical enhancements as well – if and when they become available as patented goods. In conjunction with the original goals of the Health Impact Fund (HIF), I argue this would reduce the chances that emerging enhancements will be primarily available to the wealthy, or that they will draw scarce resources away from disease treatment.

My case for the Well-Being Impact Fund (WIF) unfolds as such: First, I explore the concerns about enhancement and distribution to highlight what exactly the WIF is designed to mitigate. Second, I examine a proposal by Alan Buchanan (2011) called The Global Institute for Justice in Innovation, designed to respond to some of these concerns – I argue it is unlikely to resolve the issues I introduce below. Third, I introduce the HIF and argue that transmedical enhancements may well exacerbate the current problems facing the development and distribution of medical innovations, which I elaborate on in section three. Fourth, I show how the HIF's focus on treatments means it cannot adequately respond to the likelihood of emerging enhancements being primarily available to the wealthy. I argue the HIF should adopt a broader goal that does not distinguish between therapy and enhancement, but instead cashes out the value of both types of interventions in terms of well-being impact (that is, it ought to adopt the welfarist account of enhancement). Fourth, I examine how such a broader goal would effectively transform the HIF into a WIF and explore some hypothetical scenarios to illustrate how the latter would function. Finally, I consider a number of objections to this proposal and argue for how the WIF can overcome them.

1. Enhancement and inequality

The enhancement literature is riddled with worries about a dystopian future. The essence of these worries relates to distributive justice: if emerging enhancements are not made available to all, the argument goes, a repugnant world of oppressive inequalities will emerge. Either a minority such as the wealthy will have access to enhancements, using their superior abilities to entrench their status as an elite overclass, one that is possibly genetically distinct from the rest (e.g. Fukuyama, 2002; Mehlman, 2003; Resnik, 1994). Or, a majority will (eventually) have access to enhancements, leaving an unenhanced minority potentially excluded from, if not oppressed by, the new social and economic frameworks that will arise from having an enhanced populace (Buchanan, 2011; Wolbring, 2006). Such an unenhanced underclass would be similar to the illiterate in today's societies, if not the developmentally disabled. Either way, the disadvantage

already suffered by the poor and various minority groups would be exacerbated (Buchanan et al., 2000).

Linked to these concerns is the belief that enhancements are likely to be highly profitable to its innovators – possibly far more profitable than developing treatments for diseases. This may result in vital medical technologies being increasingly neglected in favour of new or emerging transmedical enhancements, at least in respect to research and development (Sandberg & Savulescu, 2011). If so, this would be another distributive justice problem.

To reiterate the core concerns more plainly: enhancements may exacerbate existing inequalities in wealth and liberties, and developing them may increase the neglect of disease treatment.¹ In the previous chapter, I argued that a utilitarian approach to developing and distributing enhancements would not lead to this. The loss in overall well-being from neglecting disease treatment is likely to far outweigh any increase in well-being from transmedical enhancers. Similarly, a society with an oppressive enhanced ruling elite, or an oppressed unenhanced minority, would not be one that maximizes overall well-being.

In section three, I argue these core concerns mirror existing problems facing how we develop and distribute pharmaceuticals globally, which the HIF is designed to address. Once we highlight these problems, it will become clear why it is reasonable to believe transmedical enhancements could exacerbate them in ways the HIF cannot address. As I said, I intend to tweak the HIF's goals and propose a broader Well-Being Impact Fund that can reduce the chances of this happening.

First, let us explore Allan Buchanan's (2011) Global Institute for Justice in Innovation (GIJI).

2. The Global Institute for Justice in Innovation

Buchanan (2011) calls for an institution that would work to promote wider and faster global diffusion of innovations in order to ameliorate extreme deprivation and reduce their negative impact on basic political and economic inequalities. Hence, this is not a response to the distributive issues posed by enhancements *per se*, but by all innovations. Buchanan argues the issues raised by enhancements are not necessarily unique, but fall under a wider umbrella of potentially justice diminishing innovations; hence the broader goal of his proposal.

¹ Note the specific nature of this concern: it is not a concern about the well-being of people who use these interventions, it is a concern about the oppressive or unfair consequences of those interventions for others. That is, it is not a concern that the welfarist account, on its own, engages with.

Here Buchanan (2011) sidesteps the difficulty of formulating an uncontroversial conception of justice by using an overlapping consensus approach and adopting a minimal standard: justice in this context is defined as combating undeserved and unchosen “extreme deprivation” – i.e. the severe harms that basic human rights strive to protect.

This, Buchanan (2011) argues, is one point that all theories of justice converge on as an uncontroversial goal of justice. However, he also adds combating “basic political and economic inequalities” as part of this minimal conception of justice. These refer to deeply unjust inequalities in political power and lack of access to important sites and forms of social cooperation that is of comparable consequence to the exclusion suffered by persons with disabilities in societies that do not take disability rights seriously. He deems this a practical approach because it allows us to proceed with a discussion on relevant policies without waiting for a resolution of disagreements surrounding justice more generally.

While the GIJI’s secondary purpose will be to encourage the creation of useful, justice-promoting innovations through prizes and grants, its central objective is the amelioration of injustice resulting from the monopolistic pricing of innovations by intellectual property rights holders. Its key asset is a “licensing option”. The GIJI can obtain the right to authorize compulsory licensing (CL) on a country by country basis if a patented innovation is diffusing too slowly. Buchanan defines too slowly as innovations that are failing to realize their potential for making significant gains in promoting justice, or are exacerbating existing injustices, in the form of extreme deprivation or basic political and economic inequalities.

So how might this play out with an enhancement technology? Imagine a company earned a patent to manufacture a pill that significantly increased learning ability. Let us assume it is a welfarist enhancer. Its consumers are able to lead significantly wealthier lives, expanding their boundaries of knowledge, cooperating amongst themselves in newer and more fruitful ways, and generally having a wider range of possible life plans. Such an innovation would presumably add to already existing inequalities if used by citizens of some countries only, potentially exacerbating differences in political power and economic opportunities.

Given such a patented innovation, the GIJI would publish a set of guidelines for how fast and how widely it should diffuse through the global market. Presumably the guidelines would be determined based on how critical the innovation is to justice, as minimally conceived. If the

diffusion process is slower than required by the guidelines, the GIJI can issue (or threaten to issue) a compulsory license.

How would this work? In the case of the learning enhancement envisioned above, domestic authorities of member states could initiate administrative actions that lead to a compulsory license being issued to a company in that country that can produce the otherwise patented drug and sell it locally at an affordable price. The patent-holder would receive compensation from the GIJI (which it will get through funding from its member states) at a rate between what the company was securing through monopolistic practices and what it would likely secure on the free market.

Buchanan believes that a compulsory license would rarely need to be issued since, if the GIJI is credible, just the threat of doing so would encourage patent-holders to find ways to increase the diffusion of their products.

One question this raises is whether issuing, or threatening to issue, a compulsory license for a domestic producer of some company is sufficient. It's not clear whether poorer countries (the ones that are likely to benefit most from these CLs) would have the resources to produce cutting edge innovations, whether patented or not. If the innovation is critical to justice, would the GIJI also have to subsidize the member states that make use of CLs? Alternatively, perhaps a company in a wealthier country will be allowed to manufacture the innovation and export it to the country granted a license. Either way, Buchanan might argue at this point that we ought to distinguish between inequality and injustice when it comes to patents: if a patented innovation is not diffusing fast enough even after it is sold at a lower price in a given country, it is difficult to see how the patent itself is a source of an injustice anymore.

However, the GIJI raises a number of other concerns. For instance, Hassoun (2011) argues that companies with market power are unlikely to accept the GIJI's creation to begin with. Why would they? Even if they gained some reputational benefits, as Buchanan suggests they would, Hassoun argues it is unlikely these reputational benefits will translate into any significant profits. However, Buchanan believes the GIJI offers to directly addresses concerns that producers and developed countries have with the current World Trade Organisation (WTO) sanctioned compulsory licensing procedures. In the WTO, intellectual property rights are vulnerable to compulsory licensing decisions by domestic authorities even if they only have vaguely defined

reasons. But in the GIJL, such decisions are reached within a system in which both developed countries and patent-holders themselves are active participants in the decision making process.

At this point, one may argue that surely wealthier states will use their economic and political weight, whether freely or at the behest of powerful corporate lobbies, to pressure weaker states not to exercise their authority to invoke CL. Buchanan says that there will be “rules” to control for this, such as making it illegal for one state to threaten any kind of sanction or change some trade deal if another country invokes a CL. But this seems naïve to the nature of international politics, where closed door negotiations are commonplace and officials are subject to bribes and other pressures (e.g. Lambsdorff, 1998).

This brings us to perhaps the pivotal problem with the GIJL. It suffers from the same problem most international institutions suffer from: the presumption that the domestic authorities of a country are legitimate and representative of that country’s citizenry. This presumption, which we may call international resource privilege (Pogge, 2008), refers to how any group that exercises de facto power within a country tends to be recognized internationally as the country’s legitimate government and the owner of its natural resources. Aside from the fact that this automatic recognition provides a powerful incentive toward coups and civil wars in resource rich countries, it is also clear that for a great many countries, their internationally recognized leaders are by no means representative of their citizens, often ruling autocratically and brutally, yet are nevertheless deemed the legitimate controllers of that country’s natural resources (chapter six, pp. 164).

Keeping in mind the proportion of autocratic or semi-autocratic governments in the world (Freedom House, 2016), and that these autocratic governments tend to control poorer countries – i.e. the ones presumably most likely to want to invoke CLs – how successful is Buchanan’s GIJL likely to be? Buchanan appears to want to introduce justice into a world order that is setup in a way based on an arguably unjust premise: international resource privilege.

Given this, it is no surprise that so few CLs have been invoked by poorer countries in the WTO, even though there have been many clear cases that warranted them (Hassoun, 2011). In fact, Beall and Kuhn (2012), who assembled a database of all episodes of CLs invoked since 1995, note their marked diminishment since 2006, concluding that there is a low probability of their continued usefulness.

Will the GIJI fare any better? In regards to innovations that are examples of enhancements, especially cognitive enhancements, I argue it is likely to fare worse. Imagine an autocratic or semi-autocratic government can legitimately invoke a CL for a patented cognitive enhancer which promises to increase the IQ of its citizenry by, say, an average of 5 IQ points. On the one hand, the mass use of this enhancer is likely to translate into great economic benefits, which may lend legitimacy to the autocratic government's policies. On the other, a smarter citizenry, one that is better informed, more disposed to clear, deliberative thinking, and therefore potentially more sensitive to hypocrisy and corruption, would not be what an autocratic regime seeks. Why would it invoke its right to a CL for such an innovation? More likely such a government will work to ban this enhancer from entering its market. Yet in such a scenario, the GIJI can say that it did everything it could to promote justice. And this would be true, but only because it, like most international institutions, would be premised on the acceptance of international resource privilege. Again, we can see how all this ties with my conclusion in the last chapter that we may not want to rely too much on the state for developing or distributing enhancements.

Buchanan may concede all this, but would perhaps rightly ask what any reasonable alternative might look like. Before I turn to this question, it is worth pointing out one more issue with the GIJI. Put simply, surely some innovations should be diffused at a faster rate than a market can accommodate even after a CL is issued – if, that is, they are crucial to justice. Moreover, note the issue of diversion: if a country exercises a right to a CL, there is a high risk that even if the innovation is produced cheaply there, it will be smuggled back to other countries where it is more expensive and sold at a cheaper yet still profitable price. This is one of the reasons why patent-holder resist CLs, even if they receive some compensation – which does not account for this kind of diversion (Pogge, 2005).

Despite these problems, the GIJI may have an important role to play in my proposal below. I return to it in section six of this chapter. For now, let us consider the HIF and how it can sidestep some of these challenges.

3. The Health Impact Fund

The HIF is a proposal for how to incentivize drug innovators to develop and distribute products that most effectively lessen the global burden of disease (Hollis & Pogge 2008). The implicit premise here is that drug innovators are not always motivated toward that end. Instead, profit skews drug innovation towards lucrative medical products, regardless of whether these tackle disease burden effectively or not.

To appreciate the HIF's goal, let us briefly explore how the profit-making abilities of drug innovators are currently protected. Here, the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS) is key. Put simply, TRIPS entails that World Trade Organization members must abide by 20 year-long product patents for new drugs. That is, when a company invents a new drug, it is handed a monopoly over it: no other organization can manufacture and sell it during that period. This allows it to sell its product at significant mark ups, often 50 to 100 times its cost, and in some cases 750 times (BBC, 2015).

The idea here is that allowing for such monopolies is necessary to sufficiently incentivize drug companies to take the substantial financial risks and invest the billions of dollars necessary to develop new drugs. There is little doubt this system is key to innovation, or at least to the speed with which new drugs are launched (Kyle & Qian, 2014). However, there is a difference between inventing new and useful things, and inventing new and useful things while making them affordable. One does not entail the other. Given the reward mechanism in use here, there is limited incentive to make them affordable.

And so, we end up with a number of problematic consequences to this approach to incentivizing drug innovators. Firstly, universal access to drugs is seriously compromised: it is generally more profitable for drug companies to sell their products at a price whereby only the top 15% of the human population can afford them, rather than lowering the price so a greater number can afford them (Pogge, 2015). Secondly, drug companies become more incentivized to create drugs primarily for affluent or well-insured individuals; there is comparably not much profit to be made from treating diseases that primarily affect the world's poorest (precisely because of their limited purchasing power).² Third, the current system is incredibly wasteful. Most of the money spent on pharmaceuticals – US\$1 trillion every year (Statistica, 2016; OECD, 2015) – does not go

² Pedrique et al., (2013) relay this trend emphatically: of the 850 new therapeutic products registered between 2000 and 2011, only 25 were indicated for neglected diseases. Similarly, only 1% of nearly 150,000 registered clinical trials were focused on neglected diseases.

back into manufacturing or developing drugs, but to lobbying, advertising, patenting, litigating and counterfeiting (Pogge, 2015).

These consequences, especially the first two, are evident in many of the pharmaceutical industry's decisions (Badrinath, 2014; Hays, 2015; Pedrique et al., 2013). Embodying them is this statement on a cancer treatment developed by drug company Bayer: "We did not develop this product for the Indian market, let's be honest", said by Marjin Dekkers, Bayer's then CEO. "We developed it for western patients who can afford it" (Gokhale, 2014)

This statement, along with the preceding problems arising from TRIPS and intellectual property rights more generally, suggest a crucial point: the concerns raised about enhancement and distributive injustice mirror an existing reality in regards to how we develop and distribute medicines globally. Currently, vital treatments are often sold with only the wealthy in mind, while diseases of the poor are comparatively neglected as it is not as profitable to invent affordable treatments for them.

Given this, it is reasonable to suggest that emerging enhancements could exacerbate both these trends. Transmedical enhancements need not target people with particular conditions like interventions for diseases do; anybody competent enough to understand what, say, a cognitive enhancer is could want one for themselves or children (or be convinced through marketing or social pressures they want it), but only people with, say, malaria want a malaria treatment.

This means the market for transmedical enhancement is likely to be far larger than markets for a particular treatment – something evidenced by the enormous number of prescriptions for, say, Viagra (Cockerell, 2015). This suggests that, in the same way drug companies can divert resources away from combatting deadly diseases affecting millions of poorer populations because they are not profitable, the potential profitability of transmedical enhancements may entice them to further divert those resources, and moreover to make such enhancements affordable only to wealthier demographics.³

While this possibility by no means entails a future unenhanced underclass or some elite 'genetic nobility', it is reasonable to presume enhancements will aggravate the existing problems engendered by pharmaceuticals. This is all the more so given it would likely be deemed far less

³ This is not to suggest that all transmedical enhancements will emerge through the pharmaceutical industry. The same worry applies regardless of the industry that devotes resources to developing enhancements instead of other more vital products.

controversial had Dekkers said, for instance: “We did not develop this cognitive enhancer for the poor, let’s be honest. We developed it for wealthy westerners.”

So how might the HIF help avert these problems, at least on the pharmaceuticals front? As stated its goal is to incentivize drug innovators to focus on developing products that most reduce global disease burden. More specifically, its goal is get them to sell those products at the lowest possible price, as well as ensure they reach and are used correctly by the right patients (Pogge, 2015) – all incentives that are lacking, or at least marginalized, under the current patenting system.

The essence of its mechanism is as follows: the HIF would offer drug companies the option of rewarding them over a ten-year period based on their innovation’s global health impact. In order to register, a drug innovator does not need to give up their intellectual property, but must guarantee they will offer the product at the lowest feasible cost of production and distribution (Banerjee, Hollis, & Pogge, 2010). Instead of profiting through mark-ups, they would be rewarded from a pool of funds in proportion to their product’s health impact in quality-adjusted-life-years (QALYs), which is a measure of disease burden (Vergel & Sculpher, 2008). That is, the HIF would measure the health gains the drug produces globally, and reward the company on that basis. The more the drug lessens disease burden (both, through its efficacy against a certain disease but also through how widely it is distributed), the greater its reward. In effect, the HIF is a pay-for-performance mechanism – it shifts the emphasis from sales, to actual health impact on the world (Pogge, 2015).

Crucially, the HIF would not impose this route as the only way for drug companies to profit off their products – instead, it works to supplement TRIPS, offering a purely voluntary alternative to the normal patent route.

If so, why would any drug company choose to register with the HIF? This would depend on two factors: how large and reliable the fund is⁴, and the nature of the product it has developed. If its product is directed towards affluent customers and does little in regards to health impact – say, an anti-balding remedy – then it may be more profitable for it to stay within the normal patenting system. On the other hand, if its product targets large and significant diseases concentrated amongst poor populations – say, an anti-malaria drug – the HIF offers a potentially

⁴ I will return to this point later, but the proposal is for the HIF to be largely government funded (Pogge 2015). If each participating state contributes 0.03% of its gross national income, the annual fund could start off with an initial US\$6 billion. This would allow the HIF maintain the portfolio of 20 drugs. It could also draw on a dedicated international tax and/or receive contributions from donors.

far more profitable return on its product, even and especially if the company chooses to sell that product at below cost.

With that in mind, here are some of the predicted consequences of the HIF (Pogge, 2015):

1. Prevent high prices for new, essential drug treatments, making them more easily accessible to the poor.
2. End the neglect of diseases of poverty: it incentivizes companies to develop treatments that will have the greatest health impact, rather than ones that are likely to be most profitable.
3. Drug companies would focus on products that most improve the health of the greatest number of individuals, as that is the most effective way to be rewarded under this route. As argued, the current system incentivizes selling products at prices that maximize profit regardless of whether they improve health or not.
4. Boost cost-effectiveness: it would reduce patenting costs and associated litigations, as well as reduce the need for competitive marketing. There would also be less incentives for counterfeiters as the original drugs would already be at low prices.
5. Governments and tax payers would benefit by improving healthcare efficacy and reducing the economic burdens of disease, as well as lowering the cost of insurance premiums (as drugs become cheaper to cover). Better pharmaceutical interventions are likely to reduce the reliance on hospitals and intensive care, and a less diseased populace is likely to be more economically productive.
6. Finally, companies would also gain large new markets, and new R&D opportunities, while still retaining the alternative of patenting their products or some of them.

In effect, the HIF would motivate innovators to direct resources back to where they can do the most good in terms of reducing disease burden, while also reducing inequality in terms of access to these treatments. While it doesn't promise to solve all the problems associated with how we produce and distribute medicines, it may clearly improve the status quo. Specifically, note how it side-steps the problem of international resource privilege facing the GIJ: the HIF does not need to rely on state authorities seeking to invoke a CL, nor is it susceptible to the risk of poorer or less globally dominant countries being pressured into not invoking CLs. Instead, the HIF relies on motivating innovators themselves to reduce the price of their products and to distribute them as widely as possible. Such lower prices for essential treatments would also reduce the likelihood of diversion that arises with CLs. If innovators are aiming to maximize the distribution of their products by making it universally affordable, it is unlikely to be profitable to smuggle

these treatments back to countries where these products would have otherwise been more expensive.

4. What about transmedical enhancements?

We can see how the HIF can act to mitigate the concern that emerging enhancers may exacerbate inequality or lead to neglecting disease treatment. While I argued the market for disease treatment is likely to always be smaller than the market for transmedical enhancements, the HIF can continue to make it attractive for pharmaceutical companies to develop and diffuse treatments for the most significant diseases. This can work to reduce the chances of transmedical enhancements drawing too many resources away from developing vital treatments.⁵

However, the issue here is that the HIF rewards only on the basis of disease burden reduction. This has two consequences. First, it means that, if, say, a safe, powerful cognitive enhancer were to be developed – let's say accidentally while researching a treatment for Alzheimer's – the HIF would play no role in increasing the chances of it being universally available. As an enhancer that augments well-functioning capacities rather than treats dysfunctions, such a drug would not (at least directly) reduce global disease burden, even if it does increase well-being in other ways. The HIF therefore would play no role in incentivizing its innovators to make it affordable. This means it may be available primarily to the wealthy during its patent period, bringing to relevance the concerns about worsened inequality and oppressive societies. True, it may become affordable after that period, but it seems plausible this could also be too late: during that period, the enhanced may become further enhanced and/or enriched, and an insurmountable or substantial advantage to them may arise.

Secondly, some transmedical enhancements may be hugely valuable – that is, at least as valuable as some vital treatments – and we may want to incentivize innovators to develop them as well as ensure that they are available to all in order to maximize their benefits. Ensuring that they are available to all will be key particularly to enhancements that work through a network effect (Buchanan 2011). Recall that these are interventions that are primarily beneficial to the extent that they are used by a majority. Again, the HIF, with its focus on disease impact, would

⁵ Note I am assuming the vital treatments and transmedical enhancements discussed are both also welfarist enhancements.

play no role in incentivizing developing and distributing these particular types of network-based transmedical enhancements.

What these two issues suggest is that, if it were to accommodate transmedical enhancements, the HIF needs to do things a little differently. It would need to *simultaneously* incentivize innovators to prioritize developing and distributing treatments that lessen global disease burden most, as well as transmedical enhancements likely to augment healthy people's capacities in *similarly valuable ways*.

This broader goal has a few implications. Firstly, it would incentivize innovators to invest primarily in essential treatments for deadly diseases, as well as transmedical enhancements likely to be highly valuable (as opposed to comparatively frivolous ones like some cosmetic interventions). However, since anticipating what highly valuable transmedical enhancements are is difficult, it is likely innovators will opt for less risky investments in treatments for deadly diseases (i.e. vital enhancements). On the other hand, if a transmedical enhancement were to be developed, and was for whatever reason deemed highly valuable, then innovators can register it with this expanded HIF and be rewarded on the condition they make it universally affordable and available – thus reducing the likelihood of exacerbated inequality.

Nevertheless, it is not clear what it means to talk about transmedical enhancements that are similarly valuable to medical treatments. This suggests comparing, say, the value of treating a certain number of people for a certain disease, with augmenting a certain capacity for a certain number of people. Of course, it is possible treatments for serious diseases will always be more valuable than any given transmedical enhancement; but it is also possible they won't. Here, the utilitarian deployment of the welfarist account plays crucial role.

4.1 Incorporating the welfarist account of enhancement

For the welfarist account, vital treatments for disease and transmedical interventions can both be enhancers, so long as they tend to improve well-being. In other words, the welfarist account cashes out the value of disease and transmedical intervention through their impact on well-being. What makes certain treatments "vital" is that they are vital to well-being: without it, people die or live with debilitating conditions. In other words, vital treatments are the kinds of welfarist enhancers that, in general, very reliably improve well-being (by reducing some of the worst suffering or death).

So how might the welfarist account help us compare the value of a treatment with the value of augmenting or adding a new capacity. Put simply, a treatment for a dysfunction or an augmentation of a capacity would be considered similarly valuable for a person to the extent they similarly increase their chances of leading a good life. For instance, if treating your lower back pain and augmenting your self-control would be equally advantageous to you all things considered, then they are similarly valuable – that is, they are similarly enhancing.

However, as noted in the previous chapter, this is only the part of the story covering prudential value. For overall value, we need to also factor in whatever other moral value (or disvalue) the enhancer is likely to entail. From a utilitarian perspective (although arguably from a host of other ethical perspectives), when it comes to identifying bodily and mental changes in order to prioritize them at a societal or global level (for instance, by researching interventions that cause them or by publicly funding their availability), it is this overall value that is relevant.

To elaborate, clearly some therapies and transmedical interventions may be more or less valuable overall depending on the broader effects they have that go beyond the benefits to the person undergoing them. This links back to the relationship between prudential and moral value. For instance, augmenting a person's strength may be a benefit for them but it may harm others if they use that strength to successfully mug others. In contrast, people treated for an infectious disease are personally benefited, but so are others who now will not be infected by them. Similarly, augmenting a person's cognition may benefit them personally, but it may also benefit others if they use their cognition to come up with useful tools or ideas. On the other hand, treating someone who suffers from a rare disease may benefit them personally, but harm others if doing so means underfunding other treatments for more prevalent diseases.

Again, what this interplay suggests is that, when it comes to comparing the overall value of different interventions, we ought to take into account, both, the prudential and moral value of an intervention. In the context of maximizing overall well-being through enhancement, our focus would be on a particular subset of welfarist enhancements, namely, ones that: change the biology or psychology of a person in ways that are likely to maximize overall well-being.

5. The Well-Being Impact Fund

The question now is how this ties back to the HIF. We can express the HIF's goal, as originally formulated by Hollis and Pogge (2008), as: To incentivize innovators to develop and distribute treatments in a way that most reduces global disease burden.

To incorporate transmedical enhancements to this goal, I have argued for focusing more broadly on well-being improving changes to bodies and/or minds, thus expanding our aims beyond reducing disease. However, our goal is also to prioritize those bodily and mental changes that will *most* improve overall well-being, as opposed to ones that merely contribute to overall well-being (otherwise, we have no reason to prioritize cancer treatments over anti-balding treatments, since the latter may well improve well-being too). Put together, the effect is a less disease-focused HIF, effectively transforming it into the Well-Being Impact Fund (WIF).

Its goal is: To incentivize innovators to develop and distribute interventions that change bodies and/or minds in ways that are likely to most increase overall well-being.

In this way, we have identified a particular subset of welfarist enhancement that is less individualist but that is also maximalist – in effect, we have used a utilitarian approach to deploying the account.

Such an expanded goal would have two implications. Firstly, it would reduce the chances that transmedical enhancements will draw resources away from more vital medical technologies. The reason for this is because vital medical technologies are, as the wording implies, vital to well-being. Despite a larger and wealthier market willing to pay for biomedical enhancements, the WIF can continue to make developing vital medical technologies for relatively smaller and less wealthy markets profitable through its reward mechanism.

At the same time, it may be that we can identify some transmedical interventions as incredibly valuable (in the utilitarianized sense I proposed) before they are even developed. If innovators can identify those, the WIF would incentivize investing in their development. More importantly, if a highly valuable transmedical enhancement becomes available (that is, one likely to have a strong positive impact on overall well-being), then registering it with the WIF will help increase the chances it will be affordable to most or all from day one, reducing the chances that emerging transmedical interventions will exacerbate inequalities in the ways critics of enhancement are concerned.

In this way, we can see how the WIF undertakes a broadly utilitarian task (maximizing overall well-being through bodily and mental changes to people), that also works to promote broadly egalitarian outcomes at a global level. Of course, one does not need to agree with utilitarianism to support it: the WIF is not about institutionalizing the global impartial maximization of well-being, but about using features of utilitarianism (such as aiming to improve well-being efficiently) through changes to bodies and minds. One may hold views completely contrary to utilitarianism while supporting such an approach – indeed, as we will see, the HIF itself can be accused of certain utilitarian implications, yet Pogge himself is no utilitarian (Stapleton et al., 2014).

Ultimately, it is worth immediately conceding that such a mechanism, even if highly effective and well-funded, would not eliminate the possibility of a small unenhanced underclass: some individuals may not be interested in transmedical enhancement, even if they are affordable. In this case, the WIF at least affords these individuals equality in terms of opportunities to undergo such interventions if they wish, thus mitigating the degree to which this is a concern for justice. It is also worth conceding that merely linking the value of transmedical interventions and treatments to well-being does not mean it will be easy to assess the impact of either intervention on overall well-being. This is an issue I respond to in one of the objections below. First, it is worth exploring some hypothetical scenarios to illustrate how the WIF would work.

5.1 The Well-Being Impact Fund in action

Malaria vs ovarian cancer treatment

In this scenario, two treatments have been developed, one for malaria, the other for ovarian cancer. Both are welfarist enhancers in the sense that they tend to improve people of the relevant group or individuals in the relevant circumstances (i.e. those with malaria or ovarian cancer). Here the WIF would not work much differently than the HIF. The innovators would register their products with the WIF and make their treatments available at the lowest possible price. The WIF would then monitor the impact of each drug over a ten-year period. If the drugs are both effective, then whether one receives a larger reward than the other will depend on its health impact as measured through, say, QALYs. The treatment leading to more QALYs will receive a bigger proportion of the fund.

The HIF would stop here, as it is only concerned with health impact. The WIF is concerned with overall well-being, and would therefore try to take into account where possible broader effects. For instance, in the case of ovarian cancer, it may also monitor spill-over effects for individuals who do not make use of the drug. An example may be males who do not benefit from the drug directly, but who may derive a benefit from it as an option value for a female partner, daughter, mother, or sister (O'Brien and Gafni 1996). In the case of malaria, uninfected non-users of the drug may also benefit from it by having their chances of being infected reduced. QALYs, as normally employed, would not cover such externalities (although this is not to argue they cannot be modified to carry out such a broader task). The key point here is that the WIF would strive to take all such externalities into account where possible. Here, the complex interplay between cost-effectiveness analyses, cost-utility analyses, cost benefit analyses, and cost off-set comparisons (Knapp, 2007; Labelle & Hurley, 1992) – an issue beyond my scope here – is critical to the functioning of the WIF.

Ultimately, it may be that over the first few years, the malaria drug would have a greater impact on well-being than the ovarian cancer treatment, and its innovator would therefore receive a larger reward. However, this may then shift as fewer people suffer from malaria after a few years of affordable effective treatments, and the ovarian cancer treatment innovator may then begin receiving a relatively larger share of the fund.

Malaria vs male pattern baldness treatment

This is a simpler scenario: male pattern baldness does not kill people, and while it can be associated with increased levels of anxiety and depression (Hunt & McHale, 2005), it is difficult to imagine this loss in well-being is comparable to malarial infection. Presuming this is true (and it is a revisable empirical claim subject to new evidence), there would be no incentive here for the innovator to register their product with the WIF. They can provide their anti-balding treatment through the normal patent route and at whatever high mark up. This is unproblematic: while having a full set of hair may be advantageous, it is unlikely to confer the kind of advantage relevant to the kinds of oppressive inequalities that critics of enhancement are concerned with. An oppressed underclass of specifically bald people is doubtful.

Malaria treatment vs modafinil

There is some evidence modafinil may improve some aspects of cognition (Battleday & Brem, 2015), and perhaps gives an edge in exam settings. However, like the anti-balding treatment, given its likely relatively marginal effects (Farah et al., 2014), this is unlikely to be the sort of

enhancer that would lead to the sorts of oppressive inequalities we are concerned with (although, again, this is a revisable, empirical claim). That is, there is no reason to believe a minority with access to modafinil will have privileged access to all the best jobs, say. Because of this, there is no need to try and ensure it is affordable to all people. Moreover, its long-term safety is not clear yet, and we may find regular use diminishes whatever benefit there may be to using it short-term.

More importantly, how we measure overall well-being is crucial here. It may be that a vigilance enhancer that helps individuals in, say, exam settings simply does not improve their well-being – that is, it may not be a welfarist enhancer. However, even if it is, it may be that whatever advantage it confers to them is primarily positional – as noted in chapter five. Here, if everyone starts taking it, that competitive advantage is lost. In other words, promoting its mass use by making it affordable and widely distributed would be largely self-defeating and would not increase overall well-being. In fact, given the risk of complications or long-term effects (or even just the hassle of purchasing it and taking it), its use may well diminish overall well-being. Hence, a strength of the WIF is that it is able to take into account interventions that lend themselves to such outcomes and therefore not incentivize developing or distributing them.

Finally, modafinil's patent has now expired – there would be no benefit for the innovating company to register it with the WIF, as generic manufacturers can now provide it cheaply if there is sufficient demand for it.

Malaria vs a moderate cognitive enhancer

Imagine a cognition enhancing drug is invented called “Cognovo”: it is safe and on average improves IQ by 5 points through regular use. People who take it are less forgetful, less prone to accidents, feel more alert and report improved subjective wellbeing. Let us presume that a population wide 5-point increase in IQ is associated with these changes: it would reduce poverty by 25%, reduce males in jail by 25%, reduce high school drop outs by 28%, and reduce welfare dependency by 18%. These are based on (arguably dubious) estimates by Herrnstein & Murray (1994) for the impact of a population-level 3-point IQ increase. Research by Dickerson (2006) suggests a 5-point IQ increase is worth a 40% GDP increase. We need not accept these findings as true, but they can guide our hypothetical scenario.⁶

⁶ Note also that, this hypothetical example need not imply that I am going against my commitment to avoiding highly speculative scenarios rooted in claims of inevitable or impending technologies. We need not presume anything about the likelihood of ever developing this drug – my goal here is to merely show how the WIF would operate *if* such a drug is developed.

Here, unlike modafinil, the effects of Cognovo do not appear to be primarily positional: in fact, to reap its (projected) societal or global benefits we need to ensure broad use. In this case, to compare its value to the malaria treatment, we would need a precise metric for well-being. While I will not attempt to answer which innovator the WIF would likely give a larger reward to, the point to emphasize here is the ability of the WIF to monitor a number of relevant valuables. These may range from the number of lives directly or indirectly saved by each intervention, to how and in what ways lives are otherwise improved, to the amount of resources each intervention would directly or indirectly save. For instance, in the US alone, over 3000 people are killed and half a million are injured in distraction-related crashes (NHTSA, 2015). The WIF would need to estimate the impact of using Cognovo on such figures and attempted to track that impact. Ultimately, it may turn out that, in fact, Cognovo would save far more lives worldwide, and save up far more resources that can in turn increase, say, the funds available for the malaria treatment innovator.

The key point here is to highlight the role the WIF can play in increasing the chances of these societal benefits being realised. Imagine Cognovo cost its innovators \$2 billion to develop. With their patent, they can sell it to, say, the wealthiest interested 10% for \$10 dollars per pill (making, say, \$7 billion). This would allow the company to profit far more than selling it to the wealthiest interested 70% at \$2 per pill (making, say, \$3 billion). However, doing so would mean its network effects are forgone while highlighting the concerns about inequality. Alternatively, they can register Cognovo with the WIF, sell it at or below cost of production, and instead they would profit based on the actual, measurable well-being impact Cognovo has globally. Of course, it may turn out those projected benefits are false, or simply impossible to track, in which case the innovator may wish to revert to the patent route.⁷

Next generation smart phones

I will not say much about this case except to highlight that the WIF need not be limited to innovations that are drug-based or that emerge from the pharmaceutical or biomedical industry. In fact, the idea behind the HIF has been suggested for innovations in food production and distribution, as well as “green” technologies (Pogge 2015). In this case, we can imagine an innovator develops a sustainably sourced, next generation smart phone for which we have evidence to believe would significantly increase well-being across the world if widely distributed.

⁷ The same could apply to a moral welfarist enhancer: if we have reason to believe it will reduce violence and increase cooperation without reducing users’ well-being, we may wish to ensure it is as affordable as possible in order to encourage use. Registering with the WIF can help do that.

Here, according to the extended mind thesis, this phone may in fact act as an enhancer to the extent that its functioning can form part of a users' psychology and change or augment it. Registering it with the WIF can help ensure its benefits are spread as quickly as possible by making it as affordable and widely distributed as possible.

The WIF need not accept that the extended mind thesis is true to incorporate innovations like smart phones. We could further expand the WIF to include any innovation that improves overall well-being – whether it is a tangible technology or, say, a training or educational program. This would entail a far larger – perhaps unrealistically so – fund, and for that reason we may wish to limit its scope to clear cut bodily and mental interventions, at first at least. I return to this point below.

6. Five objections

Here, I consider and respond to the five objections I think are most pertinent to this proposal. The first of these relates to the utilitarian flavour of the WIF. We can formulate the objection as such: The WIF may neglect the worst off people on the planet, if the worst off happen to be a small population. Since it aims to reward only those innovators with interventions likely to most improve overall well-being, if the worst off happen to be a small population then improving their well-being, even if it is a drastic improvement, might not be much overall compared to improving the well-being of a much larger population that is comparatively better off.

In responding to this objection, note that even the HIF, as originally formulated, is not designed to tackle, for instance, orphan diseases: horrible but rare disease afflicting very few individuals (Peterson, Hollis, & Pogge, 2010). It is concerned with global disease burden. That is, it aims to prioritize making and distributing drugs that help the most people. There is nothing counterintuitive about this. Given limited resources, if we must choose between helping a million people or 5000, we help the million, even if the 5000 are slightly worse off. However, that does not mean the minority is ignored. If the HIF is effective, that minority will benefit indirectly: by reducing the cost of medicine, and by making people healthier, government initiatives like the Orphan Drug Act (Thorat et al., 2012) may have more resources to devote to either subsidizing or developing treatments for conditions like orphan diseases. And the same point stands for the WIF. Since this not a proposal for how to fund all innovation, there are other

mechanisms that can support such a minority, whether through government subsidies, private funding, or other programs.

A second objection relates to the possibility of the fund not being large enough. Even if it is enormous, it may still be more profitable for a company to make a highly valuable biomedical enhancement only available to the world's wealthiest. For instance, in the hypothetical case of Cognovo, if its manufacturers had decided to sell it to the top 1% for \$500 per pill, that would amount to an initial profit of \$35 billion. It is unrealistic to assume the fund would be large enough to compete with that. Here, the top 1% may use this further positional advantage from being the minority with access to Cognovo to cement its privilege, triggering our concerns about oppressive inequalities arising.

One response to this is conceding that the WIF cannot stop this from happening – it must simply find a way to convince governments to augment its pool of funds to lure such companies to register with it. An alternative is to explore the possibility I alluded to above of incorporating certain aspects of Buchanan's (2011) GIJI. Recall, its main task is to monitor the diffusion of patented innovations that have may significant implications for justice. If such an innovation is diffusing globally too slowly because of pricing, the GIJI can obtain the right to authorize compulsory licensing on a country by country basis, allowing governments to have its local generics manufacturers copy the innovation and sell it cheaply without violating Intellectual Property Rights laws. While I have argued that the GIJI would face problems related to international resource privilege that the HIF or WIF would sidestep, compulsory licenses could be used as a last resort in specific cases where an innovator with a highly valuable product is insisting on massive profits and the WIF cannot compete.

A third objection relates to the various theories about well-being – a recurring issue throughout the thesis. Here, different conceptions of the good life abound and it is not clear how the WIF can promote it while there is no global consensus on what it is. However, as noted in chapter two, the first thing to highlighted here is that it is not necessarily any more difficult to agree on well-being than on what constitutes optimal health or even disease. These are by no means determinate concepts: no less than eight different disease concepts are used in clinical reasoning and practice – yet that generally does not pose a significant problem for determining who to treat and for what. This is because these different accounts of disease tend to sufficiently converge about what disease is. The same can be said of well-being. Hedonism, desire-satisfaction, and objective list theories of well-being, while offering conflicting accounts of well-being, nevertheless converge to a large extent on what makes a life go better or worse. Consider

accomplishment, friendship, autonomy, and knowledge: hedonism and desire-satisfaction theories typically hold these as instrumentally good because, for most people, they give us pleasure or satisfy our desires. Objective list theories in turn tend to recognise the value of pleasure and typically hold informed desires often reflect what is independently good (Kahane & Savulescu, 2009).

Hence, for policy purposes, it doesn't matter which theory is right: they all, perhaps for different reasons, tend to sufficiently agree. In this context, that is all we need: we are not required to deal with the grey areas of well-being because our interest is limited to incentivizing the innovation of uncontroversially good interventions. These are ones that would clearly make people's lives go better, like not dying prematurely, like not being in chronic pain or distress, but possibly also like being more intelligent, more creative, more resilient, having better self-control – these are things most would agree on as valuable.

A variation of this objection relates to *how* valuable such things are in relation to each other. While most states and cultures may agree on what is valuable in general, they may differ in the weight they give to valuable things. Indeed, there is evidence some cultures are averse to various conceptions of happiness taken for granted in the West (Joshani & Weijers, 2013). However, there is no reason why the WIF cannot take these differences into account, accommodating cultural variations that alter the impact some interventions may have on well-being in a given setting.⁸

Ultimately, weighing up what different people value would be no different than how we decide which health states to prioritize. For instance, in formulating the metrics for QALYs, people are asked whether they value a long but unhealthy life or short but healthy life, and with what trade off. They are also asked whether they value mobility more or less than freedom from discomfort, and self-care abilities more or less than anxiety. Some measures include thousands of weightings (Vergel & Sculpher, 2008; Whitehead & Ali, 2010). The same would need to be done for an approach to well-being based on overlapping consensus. An example of doing this is the OECD's Better Life Index (OECD, 2016), which attempts to measure 11 key aspects bundled up with people's well-being such as life satisfaction, community, housing, safety, education, and so on.

⁸ This does not mean the WIF has to concede to cultural relativism. Its parameters may be in line with, say, the Universal Declaration of Human Rights. Within those limitations, it can accommodate the fact that cultures give different weight to different but uncontroversial things of value, such as health, housing, safety, and so on.

The WIF would need to track the effects of an enhancement on such aspects of well-being globally.⁹

A fourth objection relates to how the WIF could possibly track the effects of a single intervention on overall well-being – that is, on potentially 7 billion or more individuals (and far more if we take into account animal well-being)? Sonderholm (2009) has levelled this criticism against the HIF and it is worth articulating in full:

...there is a further and more practical problem of applying the metric and doing the actual field work of visiting huge, poor and often geographically isolated populations and getting an accurate overview of what the disease burden is in the area and how various drugs are contributing to its reduction. Visits of this kind must be made all over the world and on a continuous basis.

This objection is all the more formidable if we are tracking well-being, which is multi-faceted and culturally variable to a degree that health may not be. However, the response here is the same: whether we are talking about health or well-being, we can concede that no metric will yield exact measures. Errors will occur on any project of this scale. However, as Peterson, Hollis and Pogge (2010) point out regarding health impact, so long as those errors are random and not too large, their effect on incentives and on payments will be small. Assessing a drug's global health impact need only be good enough, not perfect. The goal here is not to measure with accuracy as much as it is to ensure a drug innovator's overall reward is reasonably based on their drug's worldwide impact. The same would apply for a metric rooted in an overlapping set of approaches to well-being.

A further point to concede here is that tracking those effects will clearly involve some very large scale studies and intensive global monitoring capabilities. This brings us to a final, and perhaps the most salient objection: how will such an organization be funded?

As stated in footnote 4, governments would fund the HIF through 0.03% of their gross national income, though it could also rely on a dedicated international tax and donations (Pogge, 2015). Obviously the WIF may entail a much larger fund since it is interested in a wider scope of interventions. Because of this, the HIF, as originally formulated, would likely have to prove successful and convincing first before we consider expanding it with a welfarist account that accommodates a wider range of enhancements. If so, all hinges on the success of the HIF. At the time of writing, the HIF is being piloted in Mumbai to track the efficacy of new tuberculosis drugs (Mumbai Pilot Program, 2015). The World Health Organization (2009) recognizes it as

⁹ Note also that the focus on well-being at this stage is highly inclusive: we can formulate such aspects in ways that accommodate, say, Rawlsian primary goods or Nussbaum's list of capabilities.

“promising”, and progressive political parties such as the German Social Democratic Party (2010) have endorsed it. That is, it is not merely an arm-chair proposal.

7. Conclusion

As I hope we have seen, how we currently incentivize drug innovation gives us reason to suspect that the emergence of transmedical enhancements may exacerbate existing inequalities, and may draw resources away from vital treatment due to their larger profit potential. I have suggested that a proposal such as the HIF may help reduce the chances of this occurring, but not sufficiently: if a highly valuable transmedical enhancement became available, the HIF would do nothing to help ensure its affordability or wide distribution. I have suggested the HIF can maintain its core function while overcoming this limitation by adopting a broader goal that is not rooted in disease reduction, but in well-being maximization. To do that, I argued for deploying a welfarist approach, expanding the HIF into a Well-Being Impact Fund. This would incentivize innovators to develop and distribute interventions that change bodies and/or minds in ways that are likely to most increase overall well-being. Doing so would reduce the chances that transmedical enhancements will draw resources away from more vital medical technologies (to the extent that these *are* more vital to overall well-being). On the other hand, if a highly valuable transmedical enhancement becomes available, registering it with the WIF will help increase the chances it will be immediately affordable to all or most. Together, these can help avert the likelihood that transmedical enhancement will lead to oppressive inequalities or a neglect of disease treatments. More crucially, this can help mitigate existing inequalities and the neglect of certain diseases right now.

Conclusion

This thesis started with two broad questions: How should we understand the concept of enhancement? And on what basis should we promote or restrict its pursuit? The short, general answer to both these questions has been same: on a welfarist basis. Enhancements, I have argued, are bodily or mental changes that tend to improve the well-being of a person, and we should promote or restrict them based on what will tend to maximize overall well-being. As such these two questions led to the more precise goal of examining the full implications of the welfarist account, and elucidating the utilitarian basis upon which I think enhancements should be developed and distributed.

My analysis of the welfarist account of enhancement has centered on three things: objections against the account; gaps about its implications that neither the original formulators (Savulescu 2006; Savulescu et al., 2011) nor its critics have raised or addressed; and developing the account to making it more theoretically substantive. Thus, investigating the implications of the account meant answering a number of questions: Do we need to distinguish therapy from enhancement? What about distinguishing science-based interventions from so-called conventional ones? How useful is the welfarist account if there is no consensus on well-being? Does the account side-step engaging with the ethics of enhancement, ignoring the social forces that often play a crucial role in our conceptions of a good life? Answering these questions meant responding to critics of the account, but this by no means exhausted the questions raised by it: is one enhanced if their well-being improving interventions plays into or validates racist or unjust social forces? What does a change to a person's psychology entail exactly, given the extended theory of the mind? And how does the account – with its focus on what benefits the person undergoing the change – make sense of moral enhancement, which is arguably about benefits to other people? Finally, can it help practical ethicists and policy makers think more clearly about concrete cases of putative enhancements?

The preceding paragraph included a lot of questions, but each one of them directly feeds into attempting to exhaust the most significant implications of the account, subjecting it to critical scrutiny, and ultimately setting up the groundwork for answering my second question: now that we know what we mean by enhancement, on what basis should it be developed and distributed?

My approach to answering these questions has been focused on being practical: rather than committing to particular metaethical claims or assumptions, and rather than trying to resolve

long-standing theoretical disagreements, I have opted for an approach that tries to avoid rooting concepts in hazy or contestable distinctions, and one that ultimately strives for actionable conclusions and strategies for thinking about enhancement. This entailed reasoning in broadly consequentialist terms while seeking overlapping consensuses about concepts and theories, highlighting their basic elements in ways that are likely to be affirmed by those with opposing views. Given such a pragmatic end goal, I highlighted a number of ideals I would follow throughout, whereby I would avoid highly speculative claims about the likelihood of certain technologies emerging, be clear on whether I was engaged in abstract or concrete ethics, rely on empirical evidence to inform my conclusions, and acknowledge hidden assumption where possible.

So what were my key findings? Perhaps the first crucial one is that many of the criticisms of the account seem to stem from a misunderstanding of it, where the account is misconstrued as a claim about the permissibility of enhancement rather than a definition of what it *is*. We have seen this misunderstanding motivating concerns that it neglects broader societal and collective consequences of enhancement, or that it sidesteps the ethical debate by rendering enhancement as whatever contributes to the chances of a good life.

However, what these objections miss, and what I have tried to emphasize throughout, is that, by highlighting well-being, this account does not do so at the expense of other values. It only means that these values are not part of the *definition*. Hence, while it highlights well-being as a starting point for thinking about enhanced and disabled states, how and when we decide to promote an individual's well-being can also be determined by other values such as justice, as well as whatever other values individuals or society think are relevant to ethical decision making.

This focus on well-being, however, has proved controversial, even for its original formulators: is someone enhanced if, say, they alter their appearance in order to avoid racism in their society? The original formulators of the account tried to sidestep this implication of the account by excluding social prejudice from the definition, as it pertains to disability at least – no one had raised this issue from the enhancement end of the spectrum. In doing so myself, I critiqued the accounts' formulators' reasoning, arguing that excluding social prejudice from the account backfires, distorting what it is really about: advantageous bodily and mental states, regardless of whether those states are advantageous for a justifiable reason or not. It is this focus on the prudential value of a change that disentangles it from moralized and often controversial assumptions that encumber other accounts, and it is ultimately what directs debate to where it is most useful: what is a good life prudentially-speaking, and in what ways can changes to our

bodies or minds contribute to it? It is once we answer these questions that we can raise further questions: when and in what ways should people be allowed or encouraged to improve their well-being through those changes?

Another topic the associated literature had not considered is what we mean by changes to a person's psychology in the context of the account. Here, the existing literature, in line with other conceptions of enhancements to our psychology, simply takes up the folk view that changes to our psychology must be internal. I challenged that view and argued that while the welfarist account seems to retain the duality between internal and external enhancement, the extended mind thesis, if correct, suggests that some changes in a person's psychology may take place through changes in their environment. If an external prop constitutes part of a person's cognitive processes (and therefore their psychology, taken as a whole), then changes to that prop that tend to increase their chances of leading a good life (in the relevant set of circumstances) are instances of enhancement. In other words, enhancement, as it pertains to our psychologies at least, is not limited to internal changes to our bodies or brains. Objections against certain enhancements cannot be justified purely because they are "internal" – more needs to be said about what it is that makes the fact that they are internal relevant. Ultimately, the extended mind thesis may significantly broaden the scope of the account in ways hitherto unconsidered in the literature.

However, we then encountered something that had been largely neglected in the literature about the account, and that seems to limit its scope: moral enhancement. Here I highlighted a limitation of the welfarist account (it lacks the scope to capture all instances of moral enhancement), but also a positive feature of the account (it has an important role to play in assessing moral enhancers and also offers a less contentious, indirect approach to moral enhancement). And, while we may wish to distinguish welfarist enhancement from moral enhancement, I showed how those two are likely to conflict far less often than might be assumed once we examine concrete scenarios. Crucially, I elaborated on the numerous relationships prudential and moral value can share when we discuss a given intervention – relationships that can help elucidate the precise nature of our concerns or enthusiasms about a given intervention.

Moving on, I explored the topic of so-called smart drugs – are they welfarist enhancers? Here, I finally deployed the welfarist account and examined its relevance to practical ethicists and policy makers thinking about concrete interventions relevant to enhancement. My methodological approach played a key role in this chapter as I examined the literature on smart drugs, where I

critiqued prevailing presumptions in the literature and carefully examined the empirical evidence.

What we saw here is perhaps the most crucial and novel finding about the welfarist account: it offers a highly useful framework for thinking about enhancement through these two questions: What is the relationship between a change in a bodily or mental state and well-being, and how will changing it through a given intervention tend to impact well-being? As we saw, the welfarist account allows us to cash out concerns about a given intervention through a focus on well-being, assessing it on those terms. The effect is that the concerns are immediately clearer: we may disagree about whether autonomy or authenticity have intrinsic or instrumental value, but we can agree that they play an important role in our ability to lead good, meaningful lives, based on an overlapping consensus of what those are. But we also recognize that other things play such a role, such as success and intelligence, and so we must weigh those against each other before we can reach conclusions about the overall impact of an intervention on a persons' well-being. In that way, this links back to my primary goal of formulating an account that can help us think more clearly about enhancement, one that can allow bio-conservatives and proponents of enhancement to articulate their arguments in ways that that offers a constructive way forward for all.

However, the fact that the account responds to these concerns does not mean using it will exhaust all ethical issues. An intervention being of prudential value to someone or to a group does not mean its use or development is also fair or just. In other words, an ethics of welfarist enhancements (whatever these may be) would be primarily concerned with issues of justice: when should we improve people's lives and what ways should certain enhancements be prioritized or discouraged? What issues arise when we consider investing in particular types of enhancements? In other words, on what basis should we develop and distribute enhancements?

This brings us to some of the key conclusions reached in answering the second question driving this thesis. Here, I argued that in the same way well-being can act as a common currency for cashing out different values in helping us think more clearly about enhancement, the welfare-maximizing approach of utilitarianism can similarly help resolve disputes about the development and distribution of enhancements. However, in aiming to maximize well-being, utilitarianism can seemingly endorse distributions that are profoundly unfair and that violate basic rights. This objection, I argued, is based on the wealthitarian fallacy – in reality, maximizing well-being will tend to maximize equality, and utilitarian distributions are in fact often more intuitively appealing than alternative ones. However, the fact that states, which play a central role in

distribution, are unlikely to adopt utilitarianism, raises doubts about how actionable any of the conclusions I reached here are.

Hence, in the final chapter, I proposed a mechanism, the Well-Being Impact Fund, that brought the welfarist account together with a utilitarian approach. Its aim: to incentivize innovators to develop and distribute interventions that change bodies and/or minds in ways that are likely to most increase overall well-being. This, I argued, does not promise to resolve all distributive issues that enhancement raises, but it has the potential to do two things: reduce the chances that emerging transmedical enhancements will lead to oppressive inequalities in or between society, as well as the chances that these emerging technologies will lead to neglecting more vital interventions that treat disease. More crucially, to the extent that how we currently develop innovations already has these effects, the WIF can work to mitigate existing high prices for vital enhancements and the neglect of diseases of the poor.

These are an admittedly diverse number of conclusions for a thesis to reach; however, at their heart is what I hope is a framework for more fruitfully thinking about enhancement. One way to see how these conclusions link together, and specifically how they provide such a framework, is to imagine a discussion between a proponent and an opponent of a new, transmedical cognitive intervention. The proponent believes its use is permissible and that it should be widely available while the opponent does not.

As we have seen, for the opponent, the reason for their stance cannot simply be because it is not a therapy: the distinction between what is therapeutic and transmedical is blurry at best, and even when and if it can be made, it is not clear how, in and of itself, it tells us anything about the goodness or badness of an intervention. Many assumptions must be in place for such a conclusion to follow. Similarly, the fact that the intervention is not a “conventional” one, again in and of itself, tells us nothing about permissibility: I have argued conventional and science-based interventions ultimately raise the same ethical concerns and nothing about the fact that an intervention is science-based suggests that its effects are necessarily any more transient or effortless, as some have argued. Moreover, the fact that the intervention is “internal” to a person is not sufficient on its own to tell us anything morally significant. As I have argued in chapter three, our psychologies are not just internal processes, and altering the external environment can have the same, if not greater, ethical implications.

In contrast, and for the very same reasons, the proponent cannot simply assert their stance on the basis that the intervention is transmedical, or science-based, or internal. More needs to be said about why these give rise to normative reasons in favour of the intervention.

Suppose then that the opponent, with these unhelpful boundaries out of the way, concedes that their real concern is that the intervention will reduce the authenticity of the user's life. Even then, however, the proponent might respond by asking the relevance of this. It is true there may be a multitude of ways of cashing out the significance of the opponent's concern, but one way that is immediately relevant and understandable (and regardless of whether one is committed to welfarism or not) is to argue that an inauthentic life is not a good life. In other words, the concern for the opponent can be expressed as a worry that the intervention would be disabling for users.

At this point, we have a claim that most rational people can agree is something that can at least coherently be engaged with. And so the discussion morphs into something more relevant: what is authenticity and how exactly does it reduce well-being? What is the relationship between authenticity and a good life? Is an inauthentic life necessarily a bad life, so much so that any advantages of the intervention cannot outweigh this cost? The two may still disagree here, but at least the nature of their disagreement is clear now. Specifically, they are disagreeing about the two key questions I introduced in chapter five: What is the relationship between a certain bodily or mental state and well-being, and how will changing it through that transmedical intervention tend to impact well-being?

Alternatively, perhaps the opponent concedes that the intervention is not ultimately bad for users – instead, it is the collective consequences of its use that would be bad. Here, the ways prudential and moral value interact (chapter four) helps clarify this objection: we are now talking about an enhancement that has the effect of reducing moral value. Perhaps it will lead to an increase in unjustified discrimination or oppressive social relationships of some kind. At this stage, it is clear we are talking about the consequences of its use on well-being in a society or the world. The way to make sense of this concern in this framework is through the impact of those consequences on well-being. Here, utilitarianism can act as a kind of metamorality to explain the nature of these consequences: is it that we should be expending resources on more vital enhancements? Is it that only the wealthy will have access to it, using that access to cement their privilege and thereby potentially place a cap on the level of well-being of others? On the other hand, would banning it be feasible? What might be the consequences of a black market for it? Regardless of one's view of utilitarianism as a personal morality, we can see how it provides the tools to attempt answering these questions in a context that is relevant to policy-making.

Here, the proponent may concede that, while the intervention is an enhancement, perhaps we should devote resources to making other goods more available while trying to regulate its use in a way that most reduce the chances of it diminishing overall well-being. If this discussion was about whether the WIF should fund the enhancement, the proponent may concede it should be low on its list of funded or potentially funded innovations.

Of course, the opponent may still resist this entire conversation, insisting from the onset that the intervention violates some fundamental value and is in that way inherently impermissible. Perhaps it goes against nature, entails playing god, or threatens our humanity. Similarly, the proponent may insist the intervention is necessary for progress or the next step in our evolution. Whichever it is, however, if these interlocutors want to be part of a secular debate with policy implications for a plurality of people, then these concerns still have to be translated into welfarist ones: firstly, how exactly might the intervention entail going against nature, playing god or altering our human characteristics, and how and why might doing such things be bad for people? Similarly what is “progress” and why would taking the next step in our evolution be relevant to what makes lives valuable?

Hopefully this conveys an example (an admittedly abstract one) about the kind of framework the welfarist account, along with utilitarianism, can help provide for this discussion. In the introduction to this thesis, I noted that the importance of this work is straightforward: if we are to tackle the challenges posed by certain existing and emerging technologies, we need a shared, coherent and useful understanding of enhancement. That is, if we want to answer the second question driving this thesis, we need to answer the first one.

Ultimately, while I hope I have showed that these are important questions driving the thesis, this is not to suggest I have answered them without important limitations. In seeking to reach pragmatic answers with actionable conclusions, I have limited how deeply I dug into significant theoretical issues. Future research can further illuminate the implications of the welfarist account by tying it with specific theories of well-being. What are enhancements on a hedonistic account of well-being? Would the account still be plausible under hedonism? How might perfectionism, with its emphasis on exercising one’s capacities as a human being, interact with the account? Similarly, while I opted to explore deploying the account with a utilitarian approach, as I noted the account is also compatible with other ethical theories, should these be preferred. Investigating the implications of the account through a contractualist or luck egalitarian lens would be interesting, particularly in revealing just how much convergence (or

lack thereof) these theories would have with utilitarianism in terms of their real-world implications for enhancement.

Lastly, perhaps where the most further work could go is in the WIF. While I proposed it as a concrete mechanism for tackling the question of developing and distributing enhancements, it clearly remains largely conceptual at this stage. Designing metrics for well-being, and the task of actually setting up such an organization, is one that involves philosophy, economics, sociology, politics and a great range of expertise that will likely alter the WIF's scope and goals in numerous ways. Doing this work could easily take up volumes as well as extensive field work and pilot studies.

Given the breadth of this thesis, arguably each chapter could easily make a thesis of its own. However, my aim has not been to have the last word philosophically on these issues. Instead, I think the breadth of issues covered herein reflects what it takes to engage the full spectrum of skills and areas of inquiry required by practical ethics: from often abstruse conceptual analyses – what is enhancement and what is well-being? – to relatively mundane policy-making – what strategies are actually likely to catch on, rather than look good on paper? To that end, a quote from Jonathan Wolff (2011) is fitting:

In science and in social science a researcher makes his or her name by presenting a view that others find attractive or useful and build upon. By and large the situation in philosophy is the reverse. A philosopher becomes famous by arguing for a view that is highly surprising, even to the point of being irritating, but is also resistant to easy refutation.

Fame aside, it is my hope this thesis has managed to straddle a middle-path that, while perhaps unusual in some of its conclusions, offers something useful to be built upon.

Bibliography

- Adams, F., & Aizawa, K. (2010). Defending the Bounds of Cognition. In R. Menary (Ed.), *The Extended Mind* (pp. 67–80). MIT Press.
- Aditya Sharma. (2013). Cambridge Neuropsychological Test Automated Battery. In *Encyclopedia of Autism Spectrum Disorders* (pp. 498–515). New York: Springer.
- Advokat, C. (2010). What are the cognitive effects of stimulant medications? Emphasis on adults with attention-deficit/hyperactivity disorder (ADHD). *Neuroscience & Biobehavioral Reviews*, 34, 1256–1266.
- Agar, N. (2008). *Liberal Eugenics: In Defence of Human Enhancement*. John Wiley & Sons.
- Agar, N. (2014). A question about defining moral bioenhancement. *Journal of Medical Ethics*, 40, 369–370.
- Alexander, L., & Moore, M. (2015). Deontological Ethics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2015). Retrieved from <http://plato.stanford.edu/archives/spr2015/entries/ethics-deontological/>
- Amundson, R. (2005). Disability, ideology, and quality of life. In D. Wasserman & Jerome Bickenbach (Eds.), *Quality of Life and Human Difference. Genetic Testing, Health Care and Disability* (pp. 101–120).
- Areheart, B. A. (2008). When Disability Isn't Just Right: The Entrenchment of the Medical Model of Disability and the Goldilocks Dilemma. *Indiana Law Journal*, 83, 181.
- Ariely, D. (2012). *The (Honest) Truth About Dishonesty: How We Lie to Everyone---Especially Ourselves*. New York: Harper.
- Arneson, R. (2013). Egalitarianism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2013). Retrieved from <http://plato.stanford.edu/archives/sum2013/entries/egalitarianism/>

- Aronson, E. (1969). The Theory of Cognitive Dissonance: A Current Perspective. In *Advances in Experimental Social Psychology*. Academic Press.
- Aronson, E., & Carlsmith, J. M. (1962). Performance expectancy as a determinant of actual performance. *Journal of Abnormal and Social Psychology*, 65, 178–82.
- Babcock, Q., & Byrne, T. (2000). Student Perceptions of Methylphenidate Abuse at a Public Liberal Arts College. *Journal of American College Health*, 49, 143–145.
- Badrinath, R. (2014). AstraZeneca to shut down R&D site in Bangalore. Retrieved August 10, 2016, from http://www.business-standard.com/article/companies/astrazeneca-to-shut-down-r-d-site-in-bangalore-114012901153_1.html
- Banerjee, A., Hollis, A., & Pogge, T. (2010). The Health Impact Fund: incentives for improving access to medicines. *The Lancet*, 375, 166–169.
- Baranski, J., & Pigeau, R. (1997). Self-monitoring cognitive performance during sleep deprivation: effects of modafinil, d-amphetamine and placebo. *Journal of Sleep Research*, 6, 84–91.
- Baranski, J. V., Pigeau, R., Dinich, P., & Jacobs, I. (2004a). Effects of modafinil on cognitive and meta-cognitive performance. *Human Psychopharmacology: Clinical and Experimental*, 19, 323–332.
- Baranski, J. V., Pigeau, R., Dinich, P., & Jacobs, I. (2004b). Effects of modafinil on cognitive and meta-cognitive performance. *Human Psychopharmacology: Clinical and Experimental*, 19, 323–332.
- Baron, A. S., & Banaji, M. R. (2006). The Development of Implicit Attitudes Evidence of Race Evaluations From Ages 6 and 10 and Adulthood. *Psychological Science*, 17, 53–58.
- Battleday, R. M., & Brem, A. K. (2015). Modafinil for cognitive neuroenhancement in healthy non-sleep-deprived subjects: A systematic review. *European Neuropsychopharmacology*, 25, 1865–1881.

- Batty, G. D., Deary, I. J., & Gottfredson, L. S. (2007). Premorbid (early life) IQ and Later Mortality Risk: Systematic Review. *Annals of Epidemiology*, 17, 278–288.
- Baumeister, R. (1998). The Self. In *Handbook of Social Psychology* (4th ed., pp. 680–740). New York: Oxford University Press.
- BBC. (2003). *India debates “racist” skin cream ads*. Retrieved from http://news.bbc.co.uk/2/hi/south_asia/3089495.stm
- BBC. (2015). US pharmaceutical company defends 5,000% price increase. Retrieved August 9, 2016, from <http://www.bbc.com/news/world-us-canada-34320413>
- Beall, R., & Kuhn, R. (2012). Trends in Compulsory Licensing of Pharmaceuticals Since the Doha Declaration: A Database Analysis. *PLoS Med*, 9, e1001154.
- Beck, B., & Stroop, B. (2015). A Biomedical Shortcut to (Fraudulent) Happiness? An Analysis of the Notions of Well-Being and Authenticity Underlying Objections to Mood Enhancement. In J. H. Søraker, J.-W. V. der Rijt, J. de Boer, P.-H. Wong, & P. Brey (Eds.), *Well-Being in Contemporary Society* (pp. 115–134). Springer International Publishing.
- Beglinger, L. J., Gaydos, B. L., Kareken, D. A., Tangphao-Daniels, O., Siemers, E. R., & Mohs, R. C. (2004). Neuropsychological Test Performance in Healthy Volunteers Before and After Donepezil Administration. *Journal of Psychopharmacology*, 18, 102–108.
- Beglinger, L. J., Tangphao-Daniels, O., Kareken, D. A., Zhang, L., Mohs, R., & Siemers, E. R. (2005). Neuropsychological test performance in healthy elderly volunteers before and after donepezil administration: a randomized, controlled study. *Journal of Clinical Psychopharmacology*, 25, 159–165.
- Berghmans, R., ter Meulen, R., Malizia, A., & Vos, R. (2011). Scientific, Ethical, and Social Issues in Mood Enhancement. In J. Savulescu, R. ter Meulen, & G. Kahane (Eds.), *Enhancing Human Capacities* (pp. 151–165). Blackwell Publishing Ltd.
- Björklund, F., Björnsson, G., Eriksson, J., Olinder, R. F., & Strandberg, C. (2012). Recent Work on Motivational Internalism. *Analysis*, 72, 124–137.

- Blackford, R. (2012). What is this thing called enhancement? *Talking Philosophy*. Retrieved from <http://blog.talkingphilosophy.com/?p=5663>
- Bostrom, N. (2003). Human Genetic Enhancements: A Transhumanist Perspective. *The Journal of Value Inquiry*, 37, 493–506.
- Bostrom, N., & Roache, R. (2008). Ethical issues in human enhancement. *New Waves in Applied Ethics*, 120–152.
- Bostrom, N., & Roache, R. (2010). Smart Policy: Cognitive Enhancement and the Public Interest. *Contemporary Readings in Law and Social Justice*, 2, 68.
- Bostrom, N., & Sandberg, A. (2009). Cognitive enhancement: methods, ethics, regulatory challenges. *Science and Engineering Ethics*, 15, 311–341.
- Brighouse, H., & Swift, A. (2006). Equality, priority, and positional goods. *Ethics*, 116(3), 471–497.
- Brey, P. A. E. (2012). Anticipatory Ethics for Emerging Technologies. *NanoEthics*, 6, 1–13.
- British Medical Association. (2007). *Boosting your brainpower: ethical aspects of cognitive enhancements. A discussion paper from the British Medical Association*. London.
- Brinkmann, K. (2013). Better Brains or Bitter Brains? The Ethics of Neuroenhancement. In E. Hildt & A. G. Franke (Eds.), *Cognitive Enhancement* (pp. 99–112). Springer Netherlands.
- Buchanan, A., Brock, D. W., Daniels, N., & Wikler, D. (2000). *From Chance to Choice: Genetics and Justice*. Cambridge University Press.
- Buchanan, A. (2011). *Beyond Humanity? The Ethics of Biomedical Enhancement*. Oxford ; New York: Oxford University Press.
- Buchanan, A. (2011b). Cognitive enhancement and education. *Theory and Research in Education*, 9, 145–162.

- Burks, S. V., Carpenter, J. P., Goette, L., & Rustichini, A. (2009). Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proceedings of the National Academy of Sciences of the United States of America*, 106, 7745–7750.
- Bush, S. S. (2006). Neurocognitive Enhancement: Ethical Considerations for an Emerging Subspecialty. *Applied Neuropsychology*, 13, 125–136.
- Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin*, 119, 197.
- Cakic, V. (2009). Smart drugs for cognitive enhancement: ethical and pragmatic considerations in the era of cosmetic neurology. *Journal of Medical Ethics*, 35, 611–615.
- Camporesi, S. (2014). *From Bench to Bedside, to Track & Field*. San Francisco, CA: University of California Medical Humanities Press.
- Chamberlain, S. R., Robbins, T. W., Winder-Rhodes, S., Müller, U., Sahakian, B. J., Blackwell, A. D., & Barnett, J. H. (2011). Translational approaches to frontostriatal dysfunction in attention-deficit/hyperactivity disorder using a computerized neuropsychological battery. *Biological Psychiatry*, 69, 1192–1203.
- Chang, H.-J. (2011). *23 Things They Don't Tell You about Capitalism*. London: Penguin Books.
- Chang, R. (2004). Putting together morality and well-being. In P. Baumann & M. Betzler (Eds.), *Practical Conflicts: New Philosophical Essays* (pp. 118–158). Cambridge University Press.
- Cholbi, M. (2006). Belief Attribution and the Falsification of Motive Internalism. *Philosophical Psychology*, 19, 607–616.
- Clark, A. (2003). *Natural-Born Cyborgs: Minds, Technologies, and the Future of Human Intelligence*. Oxford University Press.
- Clark, A. (2009). "Response to Jerry Fodor's Review 'Where Is My Mind?'" Review of Supersizing the Mind. Retrieved from <http://www.lrb.co.uk/v31/n06/letters>
- Clark, A., & Chalmers, D. (1998). The Extended Mind. *Analysis*, 58, 7–19.

- Cockerell, J. (2015). Number of Viagra prescriptions on the rise. Retrieved August 10, 2016, from <http://www.independent.co.uk/life-style/health-and-families/health-news/viagra-number-of-prescriptions-for-erectile-dysfunction-drugs-rises-by-more-than-a-quarter-10372318.html>
- Coenen, C., Schuijff, M., & Smits, M. (2011). The Politics of Human Enhancement and the European Union. In J. Savulescu, R. ter Meulen, & G. Kahane (Eds.), *Enhancing Human Capacities* (pp. 521–535). Blackwell Publishing Ltd.
- Collingridge, D. (1980). *The social control of technology*. Frances Pinter.
- Cooper, A. (1999). The slippery slope and technological determinism. *Princeton Journal of Bioethics*, 2, 64.
- Cooper, J. M. (2007). *Cognitive Dissonance: 50 Years of a Classic Theory*. Los Angeles: SAGE Publications Ltd.
- Coyle, E. F., Hemmert, M. K., & Coggan, A. R. (1986). Effects of detraining on cardiovascular responses to exercise: role of blood volume. *Journal of Applied Physiology*, 60, 95–99.
- Crisp, R. (2016). Well-Being. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2016). Metaphysics Research Lab, Stanford University. Retrieved from <https://plato.stanford.edu/archives/sum2016/entries/well-being/>
- Crockett, M. J. (2014). Moral bioenhancement: a neuroscientific perspective. *Journal of Medical Ethics*, 40, 370–371.
- Crockett, M. J., Clark, L., Hauser, M. D., & Robbins, T. W. (2010). Serotonin selectively influences moral judgment and behavior through effects on harm aversion. *Proceedings of the National Academy of Sciences of the United States of America*, 107, 17433–17438.
- Danaher, J. (2015). Neuroenhancement and the Extended Mind Hypothesis. Retrieved from <http://philosophicaldisquisitions.blogspot.co.nz/2015/01/neuroenhancement-and-extended-mind.html>

- Daniels, N. (2000). Normal Functioning and the Treatment-Enhancement Distinction. *Cambridge Quarterly of Healthcare Ethics*, 9, 309–322.
- Deal, M. (2007). Aversive disablism: subtle prejudice toward disabled people. *Disability & Society*, 22, 93–107.
- Deary, I. J., Batty, G. D., & Gale, C. R. (2008). Bright Children Become Enlightened Adults. *Psychological Science*, 19, 1–6.
- DeGrazia, D. (1996). *Taking Animals Seriously: Mental Life and Moral Status*. Cambridge University Press.
- DeGrazia, D. (2005). *Human Identity and Bioethics*. Cambridge: Cambridge University Press.
- DeGrazia, D. (2008). Moral Status As a Matter of Degree? *The Southern Journal of Philosophy*, 46, 181–198.
- DeGrazia, D. (2012). *Creation ethics: reproduction, genetics, and quality of life*. OUP USA.
- DeGrazia, D. (2014). Moral enhancement, freedom, and what we (should) value in moral behaviour. *Journal of Medical Ethics*, 40, 361–368.
- DeMarco, J. P., & Ford, P. J. (2014). Neuroethics and the Ethical Parity Principle. *Neuroethics*, 7, 317–325.
- Dickerson, R. E. (2006). Exponential correlation of IQ and the wealth of nations. *Intelligence*, 34, 291–295.
- Dolan, R. J. (2002). Emotion, Cognition, and Behavior. *Science*, 298, 1191–1194.
- Dolcos, F., Iordan, A. D., & Dolcos, S. (2011). Neural correlates of emotion–cognition interactions: A review of evidence from brain imaging investigations. *Journal of Cognitive Psychology*, 23, 669–694.
- Douglas, T. (2013). Moral Enhancement Via Direct Emotion Modulation: A Reply to John Harris. *Bioethics*, 27, 160–168.
- Douglas, T. (2014). Moral bioenhancement, freedom and reasoning. *Journal of Medical Ethics*, 40, 359–360.

- Dunlop, M., & Savulescu, J. (2015). Distributive justice and cognitive enhancement in lower, normal intelligence. *Monash Bioethics Review*, 32, 189–204.
- Earp, B. D., Douglas, T., & Savulescu, J. (2017). Moral Enhancement. In S. Johnson & K. Rommelfanger (Eds.), *Routledge Handbook for Neuroethics*. New York: Routledge.
- Earp, B. D., Sandberg, A., Kahane, G., & Savulescu, J. (2014). When is diminishment a form of enhancement? Rethinking the enhancement debate in biomedical ethics. *Frontiers in Systems Neuroscience*, 8. Retrieved from <http://www.ncbi.nlm.nih.gov.ezproxy.otago.ac.nz/pmc/articles/PMC3912453/>
- Eggleston, B. (2014). Act utilitarianism. In B. Eggleston & D. Miller (Eds.), *The Cambridge Companion to Utilitarianism*. New York: Cambridge University Press.
- Eichhorn, L. A. (2008). Hostile Environments, Title VII, and the ADA: The Limits of the Copy and Paste Function. *Washington Law Review*, 77, 575.
- Engelbart, D. C. (1962). Augmenting human intellect: A conceptual framework. In Packer, R. and Jordan, K. *Multimedia: From Wagner to Virtual Reality*. New York: WW Norton & Company, 64–90.
- Ereshefsky, M. (2009). Defining “health” and “disease.” *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences*, 40, 221–227.
- Farah, M. J. (2011). Overcorrecting the Neuroenhancement Discussion. *Addiction*, 106, 1190–1190.
- Farah, M. J., Illes, J., Cook-Deegan, R., Gardner, H., Kandel, E., King, P., ... Wolpe, P. R. (2004). Neurocognitive enhancement: what can we do and what should we do? *Nature Reviews Neuroscience*, 5, 421–425.
- Farah, M. J., Smith, M. E., Ilieva, I., & Hamilton, R. H. (2014). Cognitive enhancement. *Wiley Interdisciplinary Reviews: Cognitive Science*, 5, 95–103.
- Farrelly, C. (2002). Genes and Social Justice: A Rawlsian Reply to Moore. *Bioethics*, 16, 72–83.

- Faust, H. S. (2008). Should we select for genetic moral enhancement? A thought experiment using the Moralkinder (MK+) haplotype. *Theoretical Medicine and Bioethics*, 29, 397–416.
- Ferrari, A., Coenen, C., & Grunwald, A. (2012). Visions and Ethics in Current Discourse on Human Enhancement. *NanoEthics*, 6, 215–229.
- Fletcher, G. (2016). *The Philosophy of Well-Being: An Introduction*. London ; New York: Routledge.
- Fletcher, G. J., Danilovics, P., Fernandez, G., Peterson, D., & Reeder, G. D. (1986). Attributional complexity: An individual differences measure. *Journal of Personality and Social Psychology*, 51, 875.
- Flynn, J. R. (2012). *Are We Getting Smarter?: Rising IQ in the Twenty-First Century*. Cambridge ; New York: Cambridge University Press.
- Forlini, C., & Racine, E. (2012). Stakeholder perspectives and reactions to “academic” cognitive enhancement: Unsuspected meaning of ambivalence and analogies. *Public Understanding of Science*, 21, 606–625.
- Fox, R. C., & Swazey, J. P. (2008). *Observing Bioethics* (1 edition). Oxford; New York: Oxford University Press.
- Franke, A. G., Bonertz, C., Christmann, M., Huss, M., Fellgiebel, A., Hildt, E., & Lieb, K. (2011). Non-medical use of prescription stimulants and illicit use of stimulants for cognitive enhancement in pupils and students in Germany. *Pharmacopsychiatry*, 44, 60-66.
- Freedom House. (2016). *Freedom in the World 2016*. Retrieved from https://freedomhouse.org/sites/default/files/FH_FITW_Report_2016.pdf
- Freitas, R. A. (1999). *Nanomedicine, volume 1: Basic capabilities*. Georgetown, TX. Retrieved from <http://www.nanomedicine.com/NMI.htm>

- Frey, M. C., & Detterman, D. K. (2004). Scholastic Assessment or g? The Relationship Between the Scholastic Assessment Test and General Cognitive Ability. *Psychological Science*, 15, 373–378.
- Fukuyama, F. (2002). *Our Posthuman Future: Consequences of the Biotechnology Revolution*. Picador.
- German Social Democratic Party. (2010). *Germany's responsibility to fight for health in developing countries - neglected diseases, reduce child and maternal mortality and a strong Global Fund*. Retrieved from <http://dip21.bundestag.de/dip21/btd/17/021/1702135.pdf>
- Gilbert, F., & Goddard, E. (2014). Thinking Ahead Too Much: Speculative Ethics and Implantable Brain Devices. *AJOB Neuroscience*, 5, 49–51.
- Gilens, M., & Page, B. I. (2014). Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens. *Perspectives on Politics*, 12, 564–581.
- Giubilini, A., & Sanyal, S. (2016). Challenging Human Enhancement. In *The Ethics of Human Enhancement: Understanding the Debate*. Oxford University Press.
- Gokhale, K. (2014). Merck to Bristol-Myers Face More Threats on India Patents. Retrieved August 10, 2016, from <http://www.bloomberg.com/news/articles/2014-01-21/merck-to-bristol-myers-face-more-threats-on-india-patents>
- Goleman, D., Boyatzis, R., & McKee, A. (2013). *Primal Leadership, With a New Preface by the Authors: Unleashing the Power of Emotional Intelligence* (10 Anv edition). Harvard Business Review Press.
- Goodin, R. E. (1995). *Utilitarianism as a Public Philosophy*. Cambridge University Press.
- Goodman, R. (2010). Cognitive Enhancement, Cheating, and Accomplishment. *Kennedy Institute of Ethics Journal*, 20, 145–160.

- Gordon, R. (2014). Enhancement and Social Possibility. Retrieved from <http://www.theneuroethicsblog.com/2014/08/enhancement-and-social-possibility.html#more>
- Gottfredson, L. (2003). g, Jobs and Life. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen*. New York: Pergamon.
- Gottfredson, L. S. (2011). Intelligence and Social Inequality. In T. Chamorro-Premuzic, S. von Stumm, & A. Furnham (Eds.), *The Wiley-Blackwell Handbook of Individual Differences* (pp. 538–575). Wiley-Blackwell.
- Gottfredson, L. S., & Deary, I. J. (2004). Intelligence Predicts Health and Longevity, but Why? *Current Directions in Psychological Science*, 13, 1–4.
- Gow, A. J., Whiteman, M. C., Pattie, A., Whalley, L., Starr, J., & Deary, I. J. (2005). Lifetime intellectual function and satisfaction with life in old age: longitudinal cohort study. *BMJ*, 331, 141–142.
- Graham, J., Haidt, J., & Nosek, B. A. (2009). Liberals and conservatives rely on different sets of moral foundations. *Journal of Personality and Social Psychology*, 96, 1029.
- Graham, J., Meindl, P., Beall, E., Johnson, K. M., & Zhang, L. (2016). Cultural differences in moral judgment and behavior, across and within societies. *Current Opinion in Psychology*, 8, 125–130.
- Graham, J., Nosek, B. A., Haidt, J., Iyer, R., Koleva, S., & Ditto, P. H. (2011). Mapping the moral domain. *Journal of Personality and Social Psychology*, 101, 366–385.
- Grau, C., Ginhoux, R., Riera, A., Nguyen, T. L., Chauvat, H., Berg, M., ... Ruffini, G. (2014). Conscious Brain-to-Brain Communication in Humans Using Non-Invasive Technologies. *PLOS ONE*, 9, e105225.
- Gray, J. R., Braver, T. S., & Raichle, M. E. (2002). Integration of emotion and cognition in the lateral prefrontal cortex. *Proceedings of the National Academy of Sciences*, 99, 4115–4120.

- Greely, H., Sahakian, B., Harris, J., Kessler, R. C., Gazzaniga, M., Campbell, P., & Farah, M. J. (2008). Towards responsible use of cognitive-enhancing drugs by the healthy. *Nature*, 456, 702–705.
- Greene, J. (2014). *Moral Tribes Emotion, Reason and the Gap Between Us and Them*. New York: Atlantic Books Ltd.
- Greene, J., & Baron, J. (2001). Intuitions about declining marginal utility. *Journal of Behavioral Decision Making*, 14, 243–255.
- Greenwald, A. G. (1980). The totalitarian ego: Fabrication and revision of personal history. *American Psychologist*, 35, 603.
- Griffin, D. W., & Ross, L. (1991). Subjective Construal, Social Inference, and Human Misunderstanding. In Mark P. Zanna (Ed.), *Advances in Experimental Social Psychology* (Vol. Volume 24, pp. 319–359). Academic Press.
- Grunwald, A. (2010). From Speculative Nanoethics to Explorative Philosophy of Nanotechnology. *NanoEthics*, 4, 91–101.
- Habermas, J. (2003). *The Future of Human Nature* (1 edition). Cambridge, UK: Polity.
- Haidt, J., & Graham, J. (2007). When Morality Opposes Justice: Conservatives Have Moral Intuitions that Liberals may not Recognize. *Social Justice Research*, 20, 98–116.
- Hamann, S. B., Ely, T. D., Grafton, S. T., & Kilts, C. D. (1999). Amygdala activity related to enhanced memory for pleasant and aversive stimuli. *Nature Neuroscience*, 2, 289–293.
- Harris, J. (2007). *Enhancing Evolution: The Ethical Case for Making Better People* (1 edition). Princeton, NJ: Princeton University Press.
- Harris, J. (2011). Moral Enhancement and Freedom. *Bioethics*, 25, 102–111.
- Harris, J. (2014). Taking liberties with free fall. *Journal of Medical Ethics*, 40, 371–374.
- Harris, J. (2016). Moral Blindness – The Gift of the God Machine. *Neuroethics*, 9, 269–273.
- Harris, J., & Chan, S. (2008). Enhancement is good for you!: Understanding the ethics of genetic enhancement. *Gene Therapy*, 15, 338–339.

- Harris, S., Mussen, P., & Rutherford, E. (1976). Some cognitive, behavioral and personality correlates of maturity of moral judgment. *The Journal of Genetic Psychology*, 128, 123–35.
- Hartog, J., & Oosterbeek, H. (1998). Health, wealth and happiness: why pursue a higher education? *Economics of Education Review*, 17, 245–256.
- Hassoun, N. (2011). Pharmaceutical Justice. In D. K. Chatterjee (Ed.), *Encyclopedia of Global Justice* (pp. 841–845). Springer Netherlands.
- Hauskeller, M. (2013). Cognitive Enhancement – To What End? In E. Hildt & A. G. Franke (Eds.), *Cognitive Enhancement* (pp. 113–123). Springer Netherlands.
- Hays, B. (2015). Snake bite crisis: Sanofi Pasteur to stop producing anti-venom. Retrieved August 10, 2016, from http://www.upi.com/Health_News/2015/09/07/Snake-bite-crisis-Sanofi-Pasteur-to-stop-producing-anti-venom/4711441655970/
- Herrnstein, R., & Murray, C. (1994). *The Bell Curve: Intelligence and Class Structure in American Life*. New York: Free Press.
- Hildt, E. (2013). Cognitive Enhancement – A Critical Look at the Recent Debate. In E. Hildt & A. G. Franke (Eds.), *Cognitive Enhancement* (pp. 1–14). Springer Netherlands.
- Hildt, E., & Franke, A. G. (Eds.). (2013). *Cognitive Enhancement: An Interdisciplinary Perspective* (2013 edition). New York: Springer.
- Hofstadter, D. (1985). *Metamagical Themas: Questing For The Essence Of Mind And Pattern*. New York: Basic Books.
- Holland, P. C., & Gallagher, M. (1999). Amygdala circuitry in attentional and representational processes. *Trends in Cognitive Sciences*, 3, 65–73.
- Hollis, A. & Pogge, T. (2008). The health impact fund: Making new medicines accessible for all. New Haven , CT : Incentives for Global Health.
- Hooker, B. (1996). Does Moral Virtue Constitute a Benefit to the Agent? In R. Crisp (Ed.), *How Should one Live?* Oxford University Press.

- Hunt, N., & McHale, S. (2005). The psychological impact of alopecia. *British Medical Journal*, 331, 951–953.
- Ilieva, I., Boland, J., & Farah, M. J. (2013). Objective and subjective cognitive enhancing effects of mixed amphetamine salts in healthy people. *Neuropharmacology*, 64, 496–505.
- Jones, G. (2008). Are smarter groups more cooperative? Evidence from prisoner's dilemma experiments, 1959–2003. *Journal of Economic Behavior & Organization*, 68, 489–497.
- Joshanloo, M., & Weijers, D. (2013). Aversion to Happiness Across Cultures: A Review of Where and Why People are Averse to Happiness. *Journal of Happiness Studies*, 15, 717–735.
- Juengst, E. (1998). What does enhancement mean? In E. Parens (Ed.), *Enhancing Human Traits: Ethical and Social Implications*. Georgetown: Georgetown University Press.
- Kahane, G. (2011). Mastery Without Mystery: Why there is no Promethean Sin in Enhancement. *Journal of Applied Philosophy*, 28, 355–368.
- Kahane, G., & Savulescu, J. (2009). The welfarist account of disability. In K. Brownlee & A. Cureton (Eds.), *Disability and disadvantage* (pp. 15–53). Oxford University Press.
- Kahneman, D. (2011). *Thinking, Fast and Slow* (1 edition). New York: Farrar, Straus and Giroux.
- Kahneman, D., & Deaton, A. (2010). High income improves evaluation of life but not emotional well-being. *Proceedings of the National Academy of Sciences*, 107, 16489–16493.
- Kahneman, D., Fredrickson, B. L., Schreiber, C. A., & Redelmeier, D. A. (1993). When More Pain Is Preferred to Less: Adding a Better End. *Psychological Science*, 4, 401–405.
- Kanazawa, S. (2010). Why liberals and atheists are more intelligent. *Social Psychology Quarterly*, 73, 33–57.
- Kapner, D. A. (2008). *Recreational Use of Ritalin on College Campuses*. *Infofacts/Resources*. Higher Education Center for Alcohol and Other Drug Abuse and Violence Prevention. Retrieved from <http://eric.ed.gov.ezproxy.otago.ac.nz/?id=ED537616>
- Kass, L. (2002). *Life Liberty & the Defense of Dignity: The Challenge for Bioethics* (1 edition). San Francisco: Encounter Books.

- Kass, L. (2003). Ageless Bodies, Happy Souls. Retrieved December 3, 2016, from <http://www.thenewatlantis.com/publications/ageless-bodies-happy-souls>
- King, C. (2016). Learning Disability and the Extended Mind. *Essays in Philosophy*, 17, 38–68.
- King, M. (2016). Limits of the welfarist account of enhancement. *Manuscript in Preparation*.
- King, M., Whitaker, M., & Jones, G. (2011). Speculative Ethics: Valid Enterprise or Tragic Cul-De-Sac? In A. Rudnick (Ed.), *Bioethics in the 21st Century*. InTech. Retrieved from <http://www.intechopen.com/books/bioethics-in-the-21st-century/speculative-ethics-valid-enterprise-or-tragic-cul-de-sac>
- Kipke, R. (2013). What Is Cognitive Enhancement and Is It Justified to Point Out This Kind of Enhancement Within the Ethical Discussion? In E. Hildt & A. G. Franke (Eds.), *Cognitive Enhancement* (pp. 145–157). Springer Netherlands.
- Kjærsgaard, T. (2015). Enhancing Motivation by Use of Prescription Stimulants: The Ethics of Motivation Enhancement. *AJOB Neuroscience*, 6, 4–10.
- Knapp, M. (2007). Economic outcomes and levers: impacts for individuals and society. *International Psychogeriatrics*, 19, 483–495.
- Koch, T. (2001). Disability and difference: balancing social and physical constructions. *Journal of Medical Ethics*, 27, 370–376.
- Koelch, M., Schnoor, K., & Fegert, J. (2008). Ethical issues in psychopharmacology of children and adolescents. *Current Opinion in Psychiatry November 2008*, 21, 598–605.
- Krueger, A. B., Kahneman, D., Fischler, C., Schkade, D., Schwarz, N., & Stone, A. A. (2008). Time Use and Subjective Well-Being in France and the U.S. *Social Indicators Research*, 93, 7–18.
- Kuhn, S. (2014). Prisoner's Dilemma. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2014). Retrieved from <http://plato.stanford.edu/archives/fall2014/entries/prisoner-dilemma/>

- Kyle, M., & Qian, Y. (2014). *Intellectual Property Rights and Access to Innovation: Evidence from TRIPS* (Working Paper No. 20799). National Bureau of Economic Research. Retrieved from <http://www.nber.org/papers/w20799>
- Labelle, R. J., & Hurley, J. E. (1992). Implications of basing health-care resource allocations on cost-utility analysis in the presence of externalities. *Journal of Health Economics*, 11, 259–277.
- Lambsdorff, J. G. (1998). An empirical investigation of bribery in international trade. *The European Journal of Development Research*, 10, 40–59.
- Lazari-Radek, K. de, & Singer, P. (2014). *The Point of View of the Universe: Sidgwick and Contemporary Ethics*. Oxford University Press. Retrieved from <http://www.oxfordscholarship.com/view/10.1093/acprof:oso/9780199603695.001.0001/acprof-9780199603695>
- Ledgerwood, A., & Trope, Y. (2011). Attitudes as global and local action guides. In J. P. Forgas, J. Cooper, & W. D. Crano (Eds.), *The Psychology of Attitudes and Attitude Change* (pp. 39–58). Psychology Press.
- Levin, J. (2013). Functionalism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2013). Retrieved from <http://plato.stanford.edu/archives/fall2013/entries/functionalism/>
- Levy, N. (2007). Rethinking Neuroethics in the Light of the Extended Mind Thesis. *The American Journal of Bioethics*, 7, 3–11.
- Levy, N. (2011). Neuroethics and the Extended Mind. In J. Illes & B. Sahakian (Eds.), *Oxford Handbook of Neuroethics*. Oxford University Press. Retrieved from <http://www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199570706.001.0001/oxfordhb-9780199570706-e-018>
- Levy, N. (2012). Ecological Engineering: Reshaping Our Environments to Achieve Our Goals. *Philosophy & Technology*, 25, 589–604.

- Levy, N. (2014). Ambivalence in the Cognitive Enhancement Debate. Retrieved from http://www.theneuroethicsblog.com/2014/10/ambivalence-in-cognitive-enhancement_14.html
- Levy, N. (2016, October 13). Would it be immoral to send out a generation starship? *Aeon Ideas*. Retrieved October 13, 2016, from <https://aeon.co/ideas/would-it-be-immoral-to-send-out-a-generation-starship>
- Lin, P., & Allhoff, F. (2008). Untangling the Debate: The Ethics of Human Enhancement. *NanoEthics*, 2, 251–264.
- Lucke, J. C., Bell, S., Partridge, B., & Hall, W. D. (2011). Deflating the Neuroenhancement Bubble. *AJOB Neuroscience*, 2, 38–43.
- Lyreskog, D. M., & Nagel, S. K. (2015). Orchestrating the Powers of the Will: Understanding Motivation Enhancement Through Higher and Lower Order Volitions. *AJOB Neuroscience*, 6, 13–15.
- Macklin, R. (2003). Dignity is a useless concept. *BMJ : British Medical Journal*, 327, 1419–1420.
- MacLean, K. A., Johnson, M. W., & Griffiths, R. R. (2011). Mystical experiences occasioned by the hallucinogen psilocybin lead to increases in the personality domain of openness. *Journal of Psychopharmacology (Oxford, England)*, 25, 1453–1461.
- Maher, B. (2008). Poll results: look who's doping. *Nature News*, 452, 674–675.
- Manson, N. (2002). Formulating the Precautionary Principle: *Environmental Ethics*, 24, 263–274.
- Maslen, H. (2015). Shrewder speculation: the challenge of doing anticipatory ethics well. *The Neuroethics Blog*. Retrieved from <http://www.theneuroethicsblog.com/2015/11/shrewder-speculation-challenge-of-doing.html?spref=tw>
- Mattay, V. S., Callicott, J. H., Bertolino, A., Heaton, I., Frank, J. A., Coppola, R., ... Weinberger, D. R. (2000). Effects of Dextroamphetamine on Cognitive Performance and Cortical Activation. *NeuroImage*, 12, 268–275.

- Mazanov, J., Dunn, M., Connor, J., & Fielding, M.-L. (2013). Substance use to enhance academic performance among Australian university students. *Performance Enhancement & Health, 2*, 110–118.
- Mazar, N., Amir, O., & Ariely, D. (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research (JMR), 45*, 633–644.
- McCabe, S. E., Teter, C. J., Boyd, C. J., Knight, J. R., & Wechsler, H. (2005). Nonmedical use of prescription opioids among U.S. college students: Prevalence and correlates from a national survey. *Addictive Behaviors, 30*, 789–805.
- Mehlman, M. (2003). *Wondergenes: Genetic Enhancement and the Future of Society*. Indiana University Press.
- Mehlman, M. (2012). *Transhumanist Dreams and Dystopian Nightmares: The Promise and Peril of Genetic Engineering*. Baltimore: Johns Hopkins University Press.
- Menary, R. (2006). Attacking the Bounds of Cognition. *Philosophical Psychology, 19*, 329–344.
- Meulen, R. ter. (2015). The Moral Ambiguity of Human Enhancement. In S. Bateman, J. Gayon, S. Allouche, J. Goffette, & M. Marzano (Eds.), *Inquiring into Human Enhancement* (pp. 86–99). Palgrave Macmillan UK.
- Mill, J. S. (1863). *Utilitarianism*.
- Mnyusiwalla, A., Daar, A. S., & Singer, P. A. (2003). “Mind the gap”: science and ethics in nanotechnology. *Nanotechnology, 14*, R9.
- Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., ... Caspi, A. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences, 108*, 2693–2698.
- Mohamed, A. D. (2014a). Neuroethical issues in pharmacological cognitive enhancement. *Wiley Interdisciplinary Reviews: Cognitive Science, 5*, 533–549.
- Mohamed, A. D. (2014b). The Effects of Modafinil on Convergent and Divergent Thinking of Creativity: A Randomized Controlled Trial. *The Journal of Creative Behavior, n/a-n/a*.

- Mohamed, A. D., & Lewis, C. R. (2014). Modafinil Increases the Latency of Response in the Hayling Sentence Completion Test in Healthy Volunteers: A Randomised Controlled Trial. *PLoS ONE*, 9, e110639.
- Mohamed, A. D., & Sahakian, B. J. (2012). The ethics of elective psychopharmacology. *International Journal of Neuropsychopharmacology*, 15, 559–571.
- Moore, A. (1994). Well-being: A philosophical basis for health services. *Health Care Analysis*, 2, 207–216.
- Mulgan, T. (2011). The Future of Utilitarianism. *The Tocqueville Review/La Revue Tocqueville*, 32, 144–166.
- Mulgan, T. (2014a). *Understanding Utilitarianism*. Stocksfield U.K.: Routledge.
- Mulgan, T. (2014b). Utilitarianism and future people. In B. Eggleston & M. Dale (Eds.), *The Cambridge Companion to Utilitarianism*. New York: Cambridge University Press.
- Mumbai Pilot Program (2015). *Healthimpactfund.org*. Retrieved from <http://healthimpactfund.org/2015/12/13/mumbai-pilot-program/>
- Murray, C. (2002). IQ and Income Inequality in a Sample of Sibling Pairs from Advantaged Family Backgrounds. *American Economic Review*, 92, 339–343.
- Napier, J. L., & Luguri, J. B. (2013). Moral Mind-Sets Abstract Thinking Increases a Preference for “Individualizing” Over “Binding” Moral Foundations. *Social Psychological and Personality Science*, 4, 754–759.
- Nave, G., Camerer, C., & McCullough, M. (2015). Does Oxytocin Increase Trust in Humans? A Critical Review of Research. *Perspectives on Psychological Science*, 10, 772–789.
- NHTSA. (2015). Distracted Driving 2013. National Highway Traffic Safety Administration. Retrieved from http://www.distracted.gov/downloads/pdfs/Distracted_Driving_2013_Research_note.pdf

- Nielsen, L. (2011). The concept of nature and the enhancement technologies debate. In J. Savulescu, R. ter Meulen, & G. Kahane (Eds.), *Enhancing Human Capacities* (pp. 19–33). Blackwell Publishing Ltd.
- Nordmann, A. (2007). If and Then: A Critique of Speculative NanoEthics. *NanoEthics*, 1, 31–46.
- Nordmann, A., & Rip, A. (2009). Mind the gap revisited. *Nature Nanotechnology*, 4, 273–274.
- Nozick, R. (1974). *Anarchy, State, and Utopia* (First Edition edition). New York: Basic Books.
- Nussbaum, M. C. (2011). *Creating Capabilities: The Human Development Approach* (1st Edition edition). Cambridge, Mass: Belknap Press.
- OECD. (2015). Pharmaceutical spending trends and future challenges. In *Health at a Glance* (pp. 29–43). Organisation for Economic Co-operation and Development.
- OECD. (2016). Better Life Initiative: Measuring Well-Being and Progress. Retrieved from <http://www.oecd.org/statistics/better-life-initiative.htm>
- Ohman, A., Flykt, A., & Esteves, F. (2001). Emotion Drives Attention: Detecting the Snake in the Grass. *Journal of Experimental Psychology*, 130, 466–478.
- Olsson, E. (2012). Coherentist Theories of Epistemic Justification. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (2012th ed.). Retrieved from <http://plato.stanford.edu/archives/spr2014/entries/justep-coherence/>
- O’Brien, B., & Gafni, A. (1996). When Do the “Dollars” Make Sense? Toward a Conceptual Framework for Contingent Valuation Studies in Health Care. *Medical Decision Making*, 16, 288–299.
- Owens, D. (2007). Disenhancement. In L. Antony (Ed.), *Philosophers without Gods*. Oxford: OUP.
- Palm, E., & Hansson, S. O. (2006). The case for ethical technology assessment (eTA). *Technological Forecasting and Social Change*, 73, 543–558.
- Parens, E. (1995). The Goodness of Fragility: On the Prospect of Genetic Technologies Aimed at the Enhancement of Human Capacities. *Kennedy Institute of Ethics Journal*, 5, 141–153.

- Parens, E. (2005). Authenticity and Ambivalence: Toward Understanding the Enhancement Debate. *Hastings Center Report*, 35, 34–41.
- Parker, E. S., Cahill, L., & McGaugh, J. L. (2006). A Case of Unusual Autobiographical Remembering. *Neurocase*, 12, 35–49.
- Partridge, B. J., Bell, S. K., Lucke, J. C., Yeates, S., & Hall, W. D. (2011). Smart Drugs “As Common As Coffee”: Media Hype about Neuroenhancement. *PLoS ONE*, 6, e28416.
- Pedrique, B., Strub-Wourgaft, N., Some, C., Olliaro, P., Trouiller, P., Ford, N., ... Bradol, J.-H. (2013). The drug and vaccine landscape for neglected diseases (2000–11): a systematic assessment. *The Lancet Global Health*, 1, e371–e379.
- Pellegrino, E. (2004). Biotechnology, Human Enhancement, and the Ends of Medicine. Retrieved May 7, 2015, from <https://cbhd.org/content/biotechnology-human-enhancement-and-ends-medicine>
- Pennycook, G., Cheyne, J. A., Barr, N., Koehler, D. J., & Fugelsang, J. A. (2014). The role of analytic thinking in moral judgements and values. *Thinking & Reasoning*, 20, 188–214.
- Perlstein, W. M., Elbert, T., & Stenger, V. A. (2002). Dissociation in human prefrontal cortex of affective influences on working memory-related activity. *Proceedings of the National Academy of Sciences*, 99, 1736–1741.
- Persson, I., & Savulescu, J. (2008). The Perils of Cognitive Enhancement and the Urgent Imperative to Enhance the Moral Character of Humanity. *Journal of Applied Philosophy*, 25, 162–177.
- Persson, I., & Savulescu, J. (2012). *Unfit for the Future: The Need for Moral Enhancement* (1 edition). Oxford: Oxford University Press.
- Persson, I., & Savulescu, J. (2016a). Enharrismment: a Reply to John Harris about Moral Enhancement. *Neuroethics*, 1–3.

- Persson, I., & Savulescu, J. (2016b). Moral Bioenhancement, Freedom and Reason. *Neuroethics*, 1–6.
- Pessoa, L. (2008). On the relationship between emotion and cognition. *Nature Reviews Neuroscience*, 9, 148–158.
- Peterson, M., Hollis, A., & Pogge, T. (2010). A Critique in Need of Critique. *Public Health Ethics*, 3, 178–185.
- Pettit, P. (2002). Non-consequentialism and political philosophy. In David Schmidtz (Ed.), *Robert Nozick* (pp. 93–104). New York: Cambridge University Press.
- Pinker, S. (2007, March 19). A History of Violence. *The New Republic*. Retrieved from <http://www.newrepublic.com/article/77728/history-violence>
- Pinker, S. (2008, May 28). The Stupidity of Dignity. Retrieved October 20, 2016, from <https://newrepublic.com/article/64674/the-stupidity-dignity>
- Pinker, S. (2011). *The Better Angels of Our Nature: Why Violence Has Declined* (First Edition edition). New York: Viking Adult.
- Pogge, T. (2005). Human Rights and Global Health: A Research Program. *Metaphilosophy*, 36, 182–209.
- Pogge, T. (2008). *World Poverty and Human Rights* (2 edition). Cambridge: Polity.
- Pogge, T. (2010). A Critique of the Capabilities Approach. In H. Brighouse & I. Robeyns (Eds.), *Measuring Justice: Primary Goods and Capabilities*. Cambridge University Press.
- Pogge, T. (2015). The Health Impact Fund: a new paradigm in pharmaceutical innovation. *Contributions to Science*, 10, 23–28.
- Pustovrh, T., & Mali, F. (2013). Exploring Some Challenges of the Pharmaceutical Cognitive Enhancement Discourse: Users and Policy Recommendations. *Neuroethics*, 7, 137–158.
- Quednow, B. B. (2010). Ethics of neuroenhancement: A phantom debate. *BioSocieties*, 5, 153–156.

- Racine, E., Rubio, T. M., Chandler, J., Forlini, C., & Lucke, J. (2014). The value and pitfalls of speculation about science and technology in bioethics: the case of cognitive enhancement. *Medicine, Health Care and Philosophy*, 17, 325–337.
- Randall, D. C., Shneerson, J. M., Plaha, K. K., & File, S. E. (2003). Modafinil affects mood, but not cognitive function, in healthy young volunteers. *Human Psychopharmacology: Clinical and Experimental*, 18, 163–173.
- Rasmussen, N. (2008). *On Speed: The Many Lives of Amphetamine* (1 edition). New York: NYU Press.
- Raus, K., Focquaert, F., Schermer, M., Specker, J., & Sterckx, S. (2014). On Defining Moral Enhancement: A Clarificatory Taxonomy. *Neuroethics*, 1–11.
- Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Belknap Press of Harvard University Press.
- Rawls, J. (1987). The Idea of an Overlapping Consensus. *Oxford Journal of Legal Studies*, 7, 1–25.
- Rawls, J. (2005). *Political Liberalism* (Expanded edition). New York: Columbia University Press.
- Redelmeier, D. A., & Kahneman, D. (1996). Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures. *Pain*, 66, 3–8.
- Repantis, D., Laisney, O., & Heuser, I. (2010). Acetylcholinesterase inhibitors and memantine for neuroenhancement in healthy individuals: A systematic review. *Pharmacological Research*, 61, 473–481.
- Repantis, D., Schlattmann, P., Laisney, O., & Heuser, I. (2010). Modafinil and methylphenidate for neuroenhancement in healthy individuals: A systematic review. *Pharmacological Research*, 62, 187–206.
- Resnik, D. (1994). Debunking the Slippery Slope Argument against Human Germ-Line Gene Therapy. *Journal of Medicine and Philosophy*, 19, 23–40.
- Resnik, D. (2000). The Moral Significance of the Therapy-Enhancement Distinction in Human Genetics. *Cambridge Quarterly of Healthcare Ethics*, 9, 365–377.

- Rip, A., & Te Kulve, H. (2008). Constructive technology assessment and socio-technical scenarios. In *Presenting Futures* (pp. 49–70). Springer.
- Roache, R. (2008). Ethics, Speculation, and Values. *NanoEthics*, 2, 317–327.
- Robeyns, I. (2006). The Capability Approach in Practice. *Journal of Political Philosophy*, 14, 351–376.
- Robeyns, I. (2011). The Capability Approach. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2011). Retrieved from <http://plato.stanford.edu/archives/sum2011/entries/capability-approach/>
- Rosati, C. S. (2016). Moral Motivation. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2016). Retrieved from <http://plato.stanford.edu/archives/fall2016/entries/moral-motivation/>
- Roskies, A. (2003). Are ethical judgments intrinsically motivational? Lessons from acquired sociopathy; [1]. *Philosophical Psychology*, 16, 51–66.
- Rowe, D. C., Vesterdal, W. J., & Rodgers, J. L. (1998). Herrnstein’s syllogism: genetic and shared environmental influences on IQ, education, and income. *Intelligence*, 26, 405–423.
- Rowlands, M. (1999). *The Body in Mind: Understanding Cognitive Processes* (First Edition edition). Cambridge, U.K. ; New York: Cambridge University Press.
- Rupert, R. D. (2004). Challenges to the Hypothesis of Extended Cognition. *The Journal of Philosophy*, 101, 389–428.
- Sagan, C. (1983). Nuclear War and Climatic Catastrophe: Some Policy Implications. *Foreign Affairs*, 62, 257–292.
- Sandberg, A., & Savulescu, J. (2011). The social and economic impacts of cognitive enhancement. *Enhancing Human Capacities*, 92–112.
- Sandel, M. J. (2007). *The Case against Perfection: Ethics in the Age of Genetic Engineering* (1 edition). Cambridge, Mass: Belknap Press of Harvard University Press.

- Sanitioso, R., Kunda, Z., & Fong, G. T. (1990). Motivated recruitment of autobiographical memories. *Journal of Personality and Social Psychology*, 59, 229–241.
- Sargant, W., & Blackburn, J. M. (1936). The effect of benzedrin on intelligence scores. *The Lancet*, 228, 1385–1387.
- Sargent, M. J. (2004). Less Thought, More Punishment: Need for Cognition Predicts Support for Punitive Responses to Crime. *Personality and Social Psychology Bulletin*, 30, 1485–1493.
- Sattler, S., Mehlkop, G., Graeff, P., & Sauer, C. (2014). Evaluating the drivers of and obstacles to the willingness to use cognitive enhancement drugs: the influence of drug characteristics, social environment, and personal characteristics. *Substance Abuse Treatment, Prevention, and Policy*, 9, 1–14.
- Savulescu, J. (2006). Justice, Fairness, and Enhancement. *Annals of the New York Academy of Sciences*, 1093, 321–338.
- Savulescu, J. (2011). Human liberation: Removing biological and psychological barriers to freedom. *Monash Bioethics Review*, 29, 04.1-04.18.
- Savulescu, J., & Birks, D. (2001). Bioethics: Utilitarianism. In *eLS*. John Wiley & Sons, Ltd.
- Retrieved from <http://onlinelibrary.wiley.com/doi/10.1002/9780470015902.a0005891.pub2/abstract>
- Savulescu, J., & Kahane, G. (2011). Disability: a welfarist approach. *Clinical Ethics*, 6, 45–51.
- Savulescu, J., & ter Meulen, R., Kahane, G. (2011). *Enhancing human capacities*. Chichester, West Sussex, U.K.; Malden, MA: Wiley-Blackwell.
- Savulescu, J., Sandberg, A., & Kahane, G. (2011). Well-Being and Enhancement. In J. Savulescu, R. ter Meulen, & G. Kahane (Eds.), *Enhancing Human Capacities* (pp. 1–18). Blackwell Publishing Ltd.
- Schaefer, G. O. (2015). Direct vs. Indirect Moral Enhancement. *Kennedy Institute of Ethics Journal*, 25, 261–289.

- Scheffler, S. (Ed.). (1988). *Consequentialism and Its Critics* (1 edition). Oxford ; New York: Oxford University Press.
- Schreiber, C. A., & Kahneman, D. (2000). Determinants of the remembered utility of aversive sounds. *Journal of Experimental Psychology: General*, 129, 27–42.
- Schwitzgebel, E. (2015). Belief. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2015). Retrieved from <http://plato.stanford.edu/archives/sum2015/entries/belief/>
- Scott, K. (2012). *Measuring Wellbeing: Towards Sustainability?* Routledge.
- Sen, A. (1973). *On Economic Inequality*. Clarendon Press.
- Sen, A., & Williams, B. (Eds.). (1982). *Utilitarianism and Beyond* (1st edition). Cambridge ; New York: Cambridge University Press.
- Shaw, W. (2009). The consequentialist perspective. In J. Dreier (Ed.), *Contemporary Debates in Moral Theory*. John Wiley & Sons.
- Shaw, W. (2014). Utilitarianism and the ethics of war. In B. Eggleston & M. Dale (Eds.), *The Cambridge Companion to Utilitarianism*. New York: Cambridge University Press.
- Shaw, W. H. (2016). *Utilitarianism and the Ethics of War*. Routledge.
- Shook, J. R. (2012). Neuroethics and the Possible Types of Moral Enhancement. *AJOB Neuroscience*, 3, 3–14.
- Sigelman, L. (1981). Is Ignorance Bliss? A Reconsideration of the Folk Wisdom. *Human Relations*, 34, 965–974.
- Silvers, A., Wasserman, D. T., & Mahowald, M. B. (1998). *Disability, Difference, Discrimination: Perspectives on Justice in Bioethics and Public Policy*. Rowman & Littlefield.
- Simpson, J. A., & Weiner, E. S. C. (Eds.). (1989). *The Oxford English Dictionary*. Oxford; Oxford; New York: Clarendon Press ; Oxford University Press.
- Singer, P. (2011). *Practical Ethics* (3 edition). New York: Cambridge University Press.

- Singh, I., Bard, I., & Jackson, J. (2014). Robust Resilience and Substantial Interest: A Survey of Pharmacological Cognitive Enhancement among University Students in the UK and Ireland. *PLoS ONE*, 9, e105969.
- Smart, J. J. C., & Williams, B. (1973). *Utilitarianism: For and Against*. Cambridge University Press.
- Smith, M. E., & Farah, M. J. (2011). Are Prescription Stimulants “Smart Pills”? The Epidemiology and Cognitive Neuroscience of Prescription Stimulant Use by Normal Healthy Individuals. *Psychological Bulletin September 2011*, 137, 717–741.
- Sonderholm, J. (2009). A Reform Proposal in Need of Reform: A Critique of Thomas Pogge’s Proposal for How to Incentivize Research and Development of Essential Drugs. *Public Health Ethics*, 3, 167-177.
- Sparrow, R. (2013). Queerin’ the PGD Clinic. *Journal of Medical Humanities*, 34, 177–196.
- Sparrow, R. (2014). Egalitarianism and Moral Bioenhancement. *The American Journal of Bioethics*, 14, 20–28.
- Spitzer, R. L. (2003). Can some gay men and lesbians change their sexual orientation? 200 participants reporting a change from homosexual to heterosexual orientation. *Archives of Sexual Behavior*, 32, 403–417.
- Stapleton, G., Schröder-Bäck, P., Laaser, U., Meershoek, A., & Popa, D. (2014). Global health ethics: an introduction to prominent theories and relevant topics. *Global Health Action*, 7. doi:10.3402/gha.v7.23569
- Statistica. (2016, August 9). Global spending on medicines 2010-2020 | Statistic. Retrieved August 9, 2016, from <http://www.statista.com/statistics/280572/medicine-spending-worldwide/>
- Stein, M. S. (2002). Utilitarianism and the Disabled: Distribution of Resources. *Bioethics*, 16, 1–19.
- Sternberg, R. J. (2011). *Cognitive Psychology* (6 edition). Belmont, Calif: Wadsworth Publishing.
- Sumner, L. W. (1992). Two Theories of the Good. *Social Philosophy and Policy*, 9, 1.

- Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success. *Journal of Personality, 72*, 271–324.
- Tännsjö, T. (2009). Ought We to Enhance Our Cognitive Capacities? *Bioethics, 23*, 421–432.
- Teitelman, E. (2001). Off-Label Uses of Modafinil. *American Journal of Psychiatry, 158*, 1341–1341.
- Temkin, L. S. (1993). *Inequality* (1 edition). New York: Oxford University Press.
- Teter, C. J., Falone, A. E., Cranford, J. A., Boyd, C. J., & McCabe, S. E. (2010). Nonmedical use of prescription stimulants and depressed mood among college students: Frequency and routes of administration. *Journal of Substance Abuse Treatment, 38*, 292–298.
- The Economist Intelligence Unit. (2015). *Democracy Index 2015: Democracy in an age of anxiety*. Retrieved from <http://www.yabiladi.com/img/content/EIU-Democracy-Index-2015.pdf>
- Thompson, D. (2007). *Ethics at Harvard 1987-2007*. Retrieved from <http://ethics.harvard.edu/what-practical-ethics>
- Thorat, C., Xu, K., Freeman, S. N., Bonnel, R. A., Joseph, F., Phillips, M. I., & Imoisili, M. A. (2012). What the Orphan Drug Act Has Done Lately for Children With Rare Diseases: A 10-Year Analysis. *Pediatrics, 129*, 516–521.
- Titchkosky, T. (2001). Disability: A Rose by Any Other Name?“People-First” Language in Canadian Society*. *Canadian Review of Sociology/Revue Canadienne de Sociologie, 38*, 125–140.
- Tressoldi, P. E., Pederzoli, L., Bilucaglia, M., Caini, P., Fedele, P., Ferrini, A., ... Accardo, A. P. (2014). *Brain-to-Brain (Mind-to-Mind) Interaction at Distance: A Pilot Study* (SSRN Scholarly Paper No. ID 2423852). Rochester, NY: Social Science Research Network. Retrieved from <http://papers.ssrn.com/abstract=2423852>

- Turner, D. C., & Sahakian, B. J. (2008). The cognition-enhanced classroom. *Reshaping the Human Condition Exploring Human Enhancement*, 107.
- UPIAS. (1976). Fundamental Principles of Disability. London: Union of the Physically Impaired Against Segregation.
- Vallentyne, P. (1994). Infinite Utility and Temporal Neutrality. *Utilitas*, 6, 193–199.
- Van Liedekerke, L., & Lauwers, L. (1997). Sacrificing the Patrol: Utilitarianism, Future Generations and Infinity. *Economics and Philosophy*, 13, 159–174.
- Vergel, Y. B., & Sculpher, M. (2008). Quality-adjusted life years. *Practical Neurology*, 8, 175–182.
- Vrecko, S. (2013). Just How Cognitive Is “Cognitive Enhancement”? On the Significance of Emotions in University Students’ Experiences with Study Drugs. *AJOB Neuroscience*, 4, 4–12.
- Wagner, G. (2013). Leveling the Playing Field: Fairness in the Cognitive Enhancement Debate. In E. Hildt & A. G. Franke (Eds.), *Cognitive Enhancement* (pp. 217–231). Springer Netherlands.
- Wagner, N.-F., Robinson, J., & Wiebking, C. (2015). The Ethics of Neuroenhancement:: Smart Drugs, Competition and Society. *International Journal of Technoethics*, 6, 1–20.
- Waldron, J. (2011). *What are Moral Absolutes Like?* (SSRN Scholarly Paper No. ID 1906850). Rochester, NY: Social Science Research Network. Retrieved from <https://papers.ssrn.com/abstract=1906850>
- Walter, S. (2010). Cognitive extension: the parity argument, functionalism, and the mark of the cognitive. *Synthese*, 177, 285–300.
- Warren, O. J., Leff, D. R., Athanasiou, T., Kennard, C., & Darzi, A. (2009). The Neurocognitive Enhancement of Surgeons: An Ethical Perspective. *Journal of Surgical Research*, 152, 167–172.
- Watanabe, M. (1990). Prefrontal unit activity during associative learning in the monkey. *Experimental Brain Research*, 80, 296–309.

- Watanabe, M. (1996). Reward expectancy in primate prefrontal neurons. *Nature*, 382, 629–632.
- West, R. F., Meserve, R. J., & Stanovich, K. E. (2012). Cognitive sophistication does not attenuate the bias blind spot. *Journal of Personality and Social Psychology*, 103, 506–519.
- Whalley, L. J., & Deary, I. J. (2001). Longitudinal cohort study of childhood IQ and survival up to age 76. *BMJ*, 322, 819.
- Whitehead, S. J., & Ali, S. (2010). Health outcomes in economic evaluation: the QALY and utilities. *British Medical Bulletin*, 96, 5–21.
- Wiegel, C., Sattler, S., Göritz, A. S., & Diewald, M. (2015). Work-related stress and cognitive enhancement among university teachers. *Anxiety, Stress, & Coping*, 0, 1–19.
- Wilson, R. A. (2004). *Boundaries of the Mind: The Individual in the Fragile Sciences - Cognition*. Cambridge University Press.
- Wilson, R. A., & Foglia, L. (2011). Embodied Cognition. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2011). Retrieved from <http://plato.stanford.edu/archives/fall2011/entries/embodied-cognition/>
- Wolbring, G. (2006). The unenhanced underclass. In *Better Humans?*
- Wolff, J. (2011). *Ethics and Public Policy: A Philosophical Inquiry* (1 edition). Milton Park, Abingdon, Oxon ; New York: Routledge.
- World Bank. (2015). Retrieved from <http://data.worldbank.org/indicator/NY.GDP.PCAP.CD>
- World Health Organization. (2009). *Public Health, Innovation and Intellectual Property: Innovation and Intellectual Property: Report of the Expert Working Group on Research and Development Financing*. Geneva.
- World Health Organization. (2015). Retrieved from <http://apps.who.int/gho/data/node.country.country-NGA?lang=en>
- Young, L., Camprodon, J. A., Hauser, M., Pascual-Leone, A., & Saxe, R. (2010). Disruption of the right temporoparietal junction with transcranial magnetic stimulation reduces the role

of beliefs in moral judgments. *Proceedings of the National Academy of Sciences*, 107, 6753–6758.

Zagorsky, J. L. (2007). Do you have to be smart to be rich? The impact of IQ on wealth, income and financial distress. *Intelligence*, 35, 489–501.

Zohny, H. (2014a). A defence of the welfarist account of enhancement. *Performance Enhancement & Health*, 3, 123–129.

Zohny, H. (2014b). Humanity Enhanced: Genetic Choice and the Challenge for Liberal Democracies by

Russell Blackford , 2013 Cambridge, MA, PB - MIT Press 248 pp., £20.95 (hb). *Journal of Applied Philosophy*, 31, 326–329.

Zohny, H. (2015). The myth of cognitive enhancement drugs. *Neuroethics*, 8, 257-269.

Zohny, H. (2016). Enhancement, disability and the riddle of the relevant circumstances. *Journal of Medical Ethics*, medethics-2015-103229.